Novel Devices and Interaction Concepts for Fluid Collaboration

Dissertation
Zur Erlangung des akademischen Grades
eines Doctor rerum naturalium (Dr. rer. nat.)

Eingereicht von
Roman Lissermann, M.Sc.
geb. in Kiew (Ukraine)

Angenommen vom
Fachbereich Informatik der Technischen Universität Darmstadt

Gutachter: Prof. Dr. Max Mühlhäuser (Technische Universität Darmstadt, DE)
Prof. Dr. Suranga Nanayakkara (Singapore University of Technology and Design, Singapore)

Tag der Einreichung: 14. Juli 2014
Tag der Disputation: 04. September 2014

D17–Darmstadt 2014

Abstract

This thesis addresses computer-augmented collaborative work. More precisely, it focuses on co-located collaboration where co-workers get together at the same place, usually a meeting room. We assume co-workers to use both mobile devices (i.e. hand-held devices) and a static device (i.e., interactive table). These devices provide multiple output modalities, such as visual output and sound output. The co-workers are assumed to process digital content (e.g., document, videos etc.).

According to both common experience and scientific evidence, co-workers often switch between rather individual, self directed work and tightly shared group work; these working styles are denoted as *loose* and *tight collaboration*, respectively. The overarching goal of this thesis is to better support seamless transitions between loose and tight collaboration, denoted as *fluid collaboration*. In order to support such fluid transitions between the two working styles, we have to reflect and mitigate conflicting requirements for both output modalities. In tight collaboration, co-workers appreciate proximity and equal access to content; both workspaces and content are shared. In loose collaboration, co-workers desire sufficient space of their own and minimal interference of their contents and interaction. It was shown that in conventional settings (e.g., interactive tables), a transition between tight and loose collaboration leads to limited personal workspace and thereby to workspace interference, clutter and other constraints. During collaboration, such interference concerns both visual and sound output.

In light of these facts, further research on interactive devices (e.g., interactive tables and mobile devices) is needed to support fluid collaboration with different output modalities. These observations lead to the central research question of this thesis: *How to support fluid co-located collaboration using visual and sound content?* This thesis explores this question in three main research directions: (1) surface-based interaction, (2) spatial interaction and (3) embodied sound interaction, while (1) and (2) address visual content, (3) focuses on auditory content. In each direction, we conceptualized, implemented, and evaluated a set of device concepts plus corresponding interaction concepts, respectively.

The first research direction, *Surface-Based Interaction*, contributes a novel tabletop, called Permulin, that provides (1) a group view providing a common ground during phases of tight collaboration, (2) private full screen views for each collaborator to scaffold loosely coupled collaboration and (3) interaction and visualization techniques for sharing content in-between these views for coordination and mutual awareness. Results from an exploratory study and from a controlled experiment provide evidence for the following advancements: (1) Permulin supports fluid collaboration by allowing the user to transition fluidly between loose and tight collaboration. (2) Users perceive and use Permulin as both a cooperative and an individual device. Amongst others, this is reflected by participants occupying significantly larger interaction areas on Permulin than on a tabletop system. (3) Permulin provides unique awareness properties: participants were highly aware of each other and of their interactions during tightly coupled collaboration, while being able to unobtrusively perform individual work during loosely coupled collaboration.

In the second research direction, *Spatial Interaction*, we simulate future paper-like display devices and investigate how well-known collaboration and interaction techniques with paper documents can be transferred to the field of video navigation based on such devices. Thereby we contribute a device concept and interaction techniques that allows multiple users to collaboratively process collections of videos on multiple paper-like displays. It enables users to navigate in video collections, create an overview of multiple videos, and structure and organize video contents. The proposed approach, coined as CoPaperVideo, leverages physical arrangement of the devices. Results of two user studies indicate that our spatial interaction concepts allow users to flexibly organize and structure multiple videos in physical space and to easily and seamlessly transition between individual and group work. In addition, the spatial interaction concepts leverage the 3D space for interaction and for mitigating space limitations.

The first two research directions contribute novel devices and interaction concepts for visual content. Visual interfaces are particularly suitable for collaboration, because they afford direct manipulation of visual content. However, while current devices support both visual and sound output, there is still a lack of suitable devices and interaction concepts for a collaborative direct manipulation of sound content. Hence, the third research direction, *Embodied Sound Interaction*, explores novel devices and interaction concepts for direct manipulation of sound for fluid collaboration. First, we contribute interfaces that enable users to control sound individually by means of *body-based interaction*. The concept focuses on the body part where sound is perceived: a user's own ear. Second, direct manipulation of sound is supported through spatial control of sound sources. Virtual sound sources are situated in 3D space and physically associated with spatially aware paper-like displays that embed videos. By physically moving these displays, each user can then control - and focus on - multiple sound sources individually or collaboratively. The evaluation supports our hypothesis that our embodied sound interaction concepts provide effective sound support for users during fluid collaboration.

Zusammenfassung

Die vorliegende Dissertation befasst sich mit der computerbasierten Zusammenarbeit mehrerer Personen. Dabei ist es üblich, dass Personen zur gemeinschaftlichen Arbeit an einem Ort einfinden. Basis dieser Arbeit ist die zunehmende Verwendung sowohl mobiler (d.h. tragbarer) als auch statischer Geräte (interaktive Tische). Diese Geräte unterstützen neben einer visuellen auch eine auditive Ausgabe, wobei digitale Inhalte, wie Dokumente, Videos und andere verwendet werden können. Wie aus der Alltagserfahrung bekannt und wissenschaftlich belegt ist, wechseln Mitarbeiter oft zwischen individueller selbstgesteuerter Arbeit und enger gemeinsamen Gruppenarbeit (im Weiteren als *lose* bzw. *enge Zusammenarbeit* bezeichnet). Diese Dissertation verfolgt das übergeordnete Ziel, computerunterstützt einen nahtlosen Übergang zwischen loser und enger Zusammenarbeit zu erlauben - diese Fähigkeit wird im Weiteren als *Fluid Collaboration* bezeichnet. Um eine solche Fluid Collaboration zu unterstützen, müssen sowohl für die visuelle als auch für die auditive Ausgabe sich entgegenstehende Anforderungen erfüllt werden. Bei *enger Zusammenarbeit* schätzen Mitarbeiter große Nähe zueinander und gemeinsamen, gleichartigen Zugang zu bearbeiteten Inhalten: sowohl der Arbeitsbereich als auch die Inhalte werden gemeinsam genutzt. Bei der *losen Zusammenarbeit* bevorzugen die Mitarbeiter ausreichend Platz als eigenes Arbeitsumfeld, eigene Inhalte und minimale Störungen bei der Interaktion. Veröffentlichungen belegen, dass der Übergang von enger zu loser Zusammenarbeit bei konventionellen Arbeitsumgebungen (Gruppenarbeits-Tisch), aber auch bei deren digitaler Entsprechung, zu starken Beschränkungen im persönlichen Arbeitsbereich führt. Als Folge davon treten Störungen bei der Interaktion, hinderliches Zusammenballen von Inhalten (Medien) und andere Einschränkungen auf. Diese Störungen betreffen sowohl die visuelle als auch auditive Ausgabe.

Angesichts dieser Tatsachen ist weitere Forschung an interaktiven Geräten (z.B. interaktive Tischen und mobilen Geräten) nötig, um die o.g. Fluid Collaboration für visuelle und auditive Ausgabe zu unterstützen. Dementsprechend lautete die zentrale Forschungsfrage der vorliegenden Dissertation: *Wie kann Fluid Collaboration unter Nutzung visueller und auditiver Inhalte im Rahmen computergestützter Zusammenarbeit an einem gemeinsamen Ort unterstützt werden?* Die Arbeit widmete sich dabei drei Hauptforschungsrichtungen, die wie nachfolgend erläutert (1) surface-based interaction, (2) spatial interaction und (3) embodied sound interaction genannt wurden. Während (1) und (2) visuelle Inhalte betreffen, konzentriert sich (3) auf die Interaktion mit auditiven Inhalten. In jeder der drei Forschungsrichtungen wurde je ein Gerätekonzept zusammen mit geeigneten Interaktionskonzepten erarbeitet, umgesetzt und ausgewertet.

In der ersten Forschungsrichtung, *Surface-Based Interaction*, wurde das Konzept eines neuen interaktiven Tisches, namens Permulin, erarbeitet. Dieser stellt folgende grundsätzliche Möglichkeiten bereit: (1) eine gemeinsame Ansicht für die Phasen der engen Zusammenarbeit, (2) eine (unter Rückgriff auf 3D-Technik realisierte) bildschirmfüllende private Ansicht für jeden Mitarbeiter für lose gekoppelte Zusammenarbeit und (3) Interaktionskonzepte und Visualisierungstechniken zum Wechsel zwischen den bei-

den Ansichten und Inhalten, um Koordination und gegenseitige Wahrnehmung zu ermöglichen. Ergebnisse aus einer explorativen Studie und einem kontrollierten Experiment zeigen dreierlei: (1) Permulin unterstützt Fluid Collaboration, indem ein nahtloser und leichter Übergang zwischen enger und loser Zusammenarbeit ermöglicht wird. (2) Benutzer nutzen Permulin sowohl sehr kooperativ als auch individuell; als Beleg dient unter anderem, dass ein signifikant größerer Interaktionsbereich bei der Benutzung von Permulin verwendet wird als bei konventionellen interaktiven Tischen. (3) Permulin verbessert erheblich die sogenannte Awareness: Teilnehmer der Studie waren sich während enger Zusammenarbeit über den wechselseitigen Status ihrer Arbeit sehr gut bewusst, konnten aber auch sehr gut ungestört in loser Zusammenarbeit arbeiten.

In der zweiten Forschungsrichtung, *Spatial Interaction*, wurden künftige so genannte papierartige Displays emuliert; es wurde untersucht, wie bekannte Interaktionstechniken und Zusammenarbeitsformen von Papierdokumenten auf die Arbeit mit computergestütztem Videomaterial übertragen werden können. Unter dem Namen CoPaperVideo entstand ein Gerätekonzept und Interaktionstechniken, womit kooperierende Benutzer auf mehreren papierähnlichen Displays in Videosammlungen navigieren und Videos (auch parallel) abspielen können. Weiterhin werden Benutzer darin unterstützt, einen räumlichen Überblick über mehrere Videos zu erstellen, Videos zu strukturieren und Videoinhalte zu organisieren. Die Ergebnisse zweier durchgeführter Benutzerstudien zeigen, dass diese räumlichen Interaktionskonzepte die Benutzer effektiv dabei unterstützen, mehrere Videos im physischen Raum flexibel zu organisieren und zu strukturieren. Dadurch wird zudem ein nahtloser und einfacher bidirektionaler Übergang zwischen Einzel- und Gruppenarbeiten realisiert. Darüber hinaus nutzt die räumliche Interaktion den dreidimensionalen Raum aus und überwindet so Beschränkungen die auf einer (2D-) Fläche auftreten.

Direkte Manipulation bietet sich für visuelle Benutzerschnittstellen, insbesondere im Zusammenhang mit Kooperation, an. Die genannten Geräte unterstützen aber sowohl die visuelle als auch auditive Ausgabe. Im Bereich auditiver Inhalte sind kaum geeignete Geräte für kollaborative und direkte Manipulation dieser zu finden. Daher befasst sich die dritte Forschungsrichtung, *Embodied Sound Interaction*, mit neuartigen Geräten und Interaktionskonzepten für die direkte Manipulation von auditiven Inhalten im SInne der Fluid Collaboration. Dabei wird im ersten Schritt direkte Manipulation mit auditiven Inhalten ermöglicht, indem Benutzer das eigene Ohr nach der so genannten On-Body-Interaction verwenden, um individuell mit auditiven Inhalten zu interagieren. Im zweiten Schritt wird eine direkte Manipulation von mehreren Tonquellen ermöglicht durch die räumliche Steuerung jeder einzelnen Tonquelle. Virtuelle Tonquellen werden dabei im 3D-Raum positioniert und durch physische Interaktion räumlich mit papierähnlichen Displays (welche die zum Audio gehörigen Videos beinhalten) manipuliert. Durch Verschieben eines der papierähnlichen Displays kann jeder Benutzer mehrere Tonquellen einzeln oder gemeinsam steuern. Die Ergebnisse der Evaluation zeigen, dass die direkten Manipulationskonzepte mit Tonquellen eine effektive Unterstützung der Benutzer während der Fluid Collaboration bieten.

Acknowledgments

I would like to acknowledge the support, patience, and guidance of the following people. Without them, this dissertation, which is so far definitely the hardest journey I ever took, would not have been completed.

Supervision: I would like to start and thank my supervisors. Their wisdom, knowledge, and commitment to the highest standards motivated and inspired me.

I would like to first thank my advisor and mentor, *Prof. Dr. Max Mühlhäuser* (Technische Universität Darmstadt, Germany). Max is a very remarkable person. Despite our short meetings, which we both enjoyed, Max has strongly and with much focus guided my dissertation always into the right direction. I would like to thank Max for his supervision, his effort and his patience during my dissertation. In addition, I would like to thank Max because in most of my questions, where I asked him for help, he always answered with yes and tried to make my ideas and requests happen. I think this is a remarkable talent.

I am also very grateful to *Prof. Dr. Suranga Nanayakkara* (Singapore University of Technology and Design, Singapore) for acting as a co-referee in this thesis. Your valuable feedback made a big change to this work.

A big thanks to *Dr. Jochen Huber* (MIT Media Lab, USA; and Singapore University of Technology and Design, Singapore), who is my colleague and, more importantly, my friend. He always supported, encouraged, and believed in me. Without his support and constant patience with me, this dissertation would have never become reality. Jochen is a wonderful person; he is always in a good mood and has an open ear to listen to your problems. Jochen is the most rooted-to-the-soil person that I have ever met. Thanks, Jochen, for being part of my work and daily life.

I would also like to thank *Dr. Jürgen Steimle* (Max Planck Institute for Informatics, Germany; and MIT Media Lab, USA). Jürgen was my first postdoc supervisor in this thesis. In my opinion, he placed the first stones of my understanding of how to conduct research and how to ask myself the right questions. Jürgen is an incredible person, who brings such a motivation to work that I have so far never seen in others. Jürgen has pushed and driven me to the highest standards, not only in research, but also in myself, which inspires me until now.

Colleagues and students: I would like to give a big thanks to my collegues and friends, who have served me with valuable discussions and support: Mo Khalilbeigi, Niloofar Dezfuli, Simon Olberding, Martin Schmitz, Jan Riemann, and Florian Müller. I would like to thank all my students. Without them, I would have never finished this work. In particular, I would like to thank: Benjamin Petry, Martin Schmitz, Matthias Eichholz, Rene Hertling, Fawaz Malik, Eric Brüggemann, Mohit Makhija, Ilja Schwarz, Marcel Dostal, Martin Distler, Simon Werner, Deborah Buth, Dominik Rohn, Jerome Möckel, Arvid Lange, Michael Stoica, Daniel Vancura, Pascal Weisenburger, Christian Brückner, Tim Wimmer, Karsten Kruse, and Manuel Krönig. Thank you all for your valuable contributions.

Family and friends: I would like to thank the whole Lissermann family. Specially my parents, Natalia and Josef Lissermann, who have chosen to leave their lives, their university degrees, and their hometown back in Ukraine to invest *everything* into my brother Grigorij Lissermann and me. Furthermore, I would like to thank, from the bottom of my heart, my closest friends Daniel, Andrej, and Ilja. Last but not least, I am deeply grateful for the love and support of Deniz in the last miles of this work.

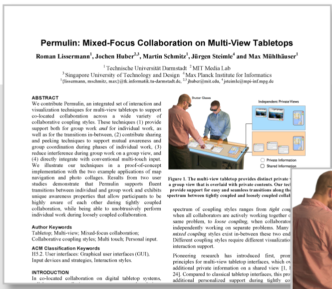
To my parents (Natalia and Josef Lissermann)
who have always loved and believed in me...

THESIS CHAPTERS | SELECTED PUBLICATIONS

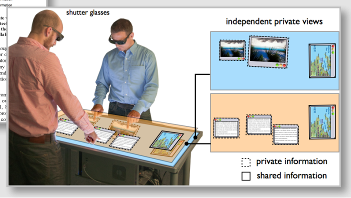
Introduction

I. SURFACE-BASED INTERACTION

2.1 Scope of Suface-Based Interaction
2.2 Related Work
2.3 Supporting Fluid Information Spaces
2.4 Interaction Concepts
2.5 Evaluation
2.6 Conclusion

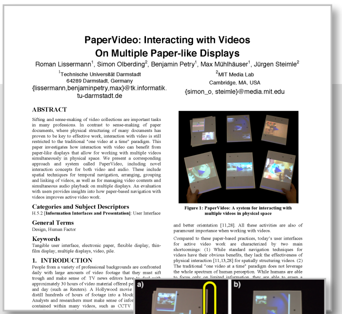


Roman Lissermann, Jochen Huber, Martin Schmitz, Jürgen Steimle, Max Mühlhäuser. *Permulin: Mixed-Focus Collaboration on Multi-View Tabletops*. In Proc. CHI '14, 3191-3200.



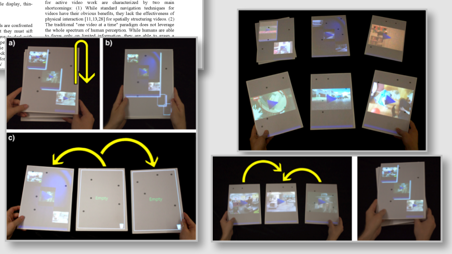
II. SPATIAL INTERACTION

3.1 Scope of Spatial Interaction
3.2 Related Work
3.3 Design Space of Multiple Spatially Aware Displays
3.4 Supporting Collaborative Spatial Interaction with Videos
3.5 Interaction Concepts with Videos on Multiple Paper-like Displays
3.6 Implementation
3.7 Evaluation
3.8 Conclusion



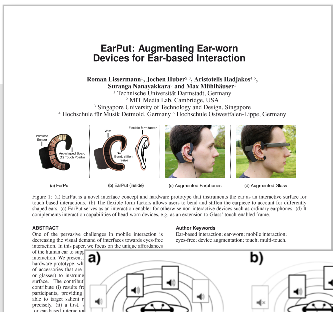
Roman Lissermann, Simon Olberding, Benjamin Petry, Max Mühlhäuser, and Jürgen Steimle. *PaperVideo: interacting with videos on multiple paper-like displays*. In Proc. MM '12, 129-138.

Roman Lissermann, Simon Olberding, Max Mühlhäuser, and Jürgen Steimle. *Interacting with Videos On Paper-like Displays*. In Proc. CHI EA '12, 2579-2584.



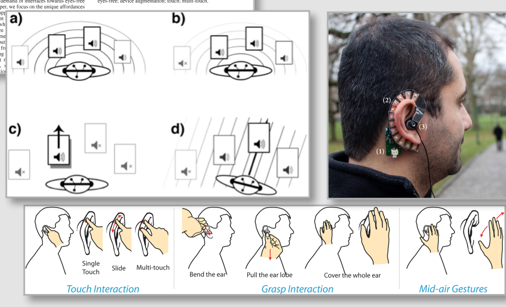
III. EMBODIED SOUND INTERACTION

4.1 Related Work
4.2 Supporting Collaborative Concept for Embodied Sound Interaction
4.3 Single-User Ear-Based Sound Control
4.4 Single-User Spatial Interaction Concepts for Parallel Sound Sources
4.5 Multi-User Spatial Interaction Concepts for Parallel Sound Sources
4.6 Conclusion



Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. *EarPut: Augmenting Ear-worn Devices for Ear-based Interaction*. In Proc. OzCHI '14, (in press).

Roman Lissermann, Simon Olberding, Benjamin Petry, Max Mühlhäuser, and Jürgen Steimle. *PaperVideo: interacting with videos on multiple paper-like displays*. In Proc. MM '12, 129-138.



Conclusion

Contents

1 Introduction

Contents

1.1 Motivation

The present thesis addresses computer support for so-called *co-located meetings*, that is, the gathering of small teams of two to six people at one location with a common goal. The goal is supported to imply the selection, consumption, manipulation, and production of information that is, at least in part, digitally represented (document, videos, etc.).

In co-located collaborations, users constantly have to switch between individual and group work. This is also called *mixed-focus collaboration* [Gutwin and Greenberg, 1998]. When *working in a group*, users usually require a shared common view to have the same point of reference for discussion. When *working individually*, users partition themselves and occupy their own spaces. Thereby, different coupling[1] styles exist that have been introduced by Tang *et al.* [2006]. The **spectrum of coupling** ranges from *tight coupling*, when all collaborators are actively working together on the same problem, to *loose coupling*, when collaborators are independently working on separate problems. This, however, requires collaborators to coordinate their interactions through, for example, partitioning the surface into dedicated personal and group territories [Scott *et al.*, 2004; Tse *et al.*, 2004]. Working on different views also requires spatial partitioning of the screen into several smaller views [Isenberg *et al.*, 2012]. Partitioning the surface however results in limited space for interaction. This space limitation increasingly occurs when users flexibly transition between tight and loose coupling in mixed-focus collaboration (further called **fluid collaboration**).

Nowadays, it has become common that meeting rooms are equipped with devices that allow support for digital content. For example, interactive tables replace conventional tables. In addition, recent technology advancements made portable personal devices ubiquitous. It has become common to bring these devices to meetings as well (c.f. BYOD: bring your own device). These mobile devices support, for example, paper documents with additional digital content.

These devices provide multiple output modalities, such as **visual output** and **sound output**. During a collaboration, interacting with and controlling these modalities for a group of people, however, lead to multiple limitations and challenges for each collaborator, because of the current hardware setting. **Interactive tables** have a constant awareness for all collaborators, leading to *limited interaction space, screen clutter, and workspace interference* for the visual output. For sound output, if sound is played back by multiple users on multiple devices, this leads to *sound clutter and overload of each user* in terms of auditive perception. Furthermore, known interaction paradigms with paper documents, for example, spatially structuring and organization for collaboration, are so far not supported with digital content. In particular, because today's **mobile devices** are *heavy, have only limited interconnection, and are spatially unaware about each other and their position* for effectively supporting spatial interaction paradigms known from paper documents, for example, piling and spatially arranging.

Thereby, interfaces are needed to support fluid collaboration while maximizing proximity, conserving *close phase social distance*[2] [Hall and Hall, 1969], (a) on interactive surfaces (surface-based interaction), (b) above interactive surfaces (spatial interaction), as well as (c) in settings with multiple output modalities (embodied sound interaction). Along these lines, this thesis addresses the following central research questions:

[1]Coupling is the amount of work a person can do before new instructions or a discussion with another person is needed (see Definition 3).

[2]Close phase social distance ranges from 1.2 to 2.1 m.

How do we design and implement novel devices and interaction concepts that support fluid collaboration in close physical proximity?

In each direction, this thesis contributes *novel devices and interaction concepts* for co-located fluid collaboration (see Figure 1.1). In the following, we briefly present each research direction.

1. **Surface-Based Interaction:** Commercially available interactive tabletops have made the first steps by providing collaborators simultaneous access to digital content in meetings. During a mixed-focus collaboration, however, users are forced to partition their screen space, which leads to a *limited interaction space* on tabletops.

   In order to bridge this limitation and support individual work and group work on such interactive surfaces, each user needs a private as well as a group view of the very same surface. To realize this, novel devices and interaction concepts are required and will be presented in this thesis.

2. **Spatial Interaction:** With regard to collaborative space management, research shows that the key is using not only one but multiple documents or sheets of paper simultaneously, in order to manipulate and organize information in physical space. Furthermore, it has been proven to effectively support comparison, overview generation and better orientation [Kirsh, 1995; Sellen and Harper, 2001]. These activities are also important during fluid collaboration. As it has become common to bring your own mobile devices to meetings, today these devices are too heavy, only have limited interconnection, and are spatially unaware about each other to support the previously stated organizing and structuring activities.

   Given the rapid advances in mobile devices, future tablet devices are very likely to be flexible, thin, and lightweight compared with contemporary mobile devices [Co and Pashenkov, 2008; Crawford, 2005]. This, further called *paper-like displays*, will eventually lead to another type of device brought to meetings specifically for collaboration. These devices will be very similar to paper so that multiple paper-like displays could allow users to bring their digital content to meetings. Digital content on such paper-like displays could be laid out in space, piled, and passed over to others like paper. This *spatial interaction* that leverages 3-D space for interaction could also bridge space limitation and allow a new way of space management for fluid collaboration.

3. **Embodied Sound Interaction:** Previously mentioned two research directions focus on interactive tabletops and multiple paper-like displays incorporate the visual modality of users during fluid collaboration. Visual interfaces are particularly suitable for collaboration because of their support of direct manipulation [Hutchins

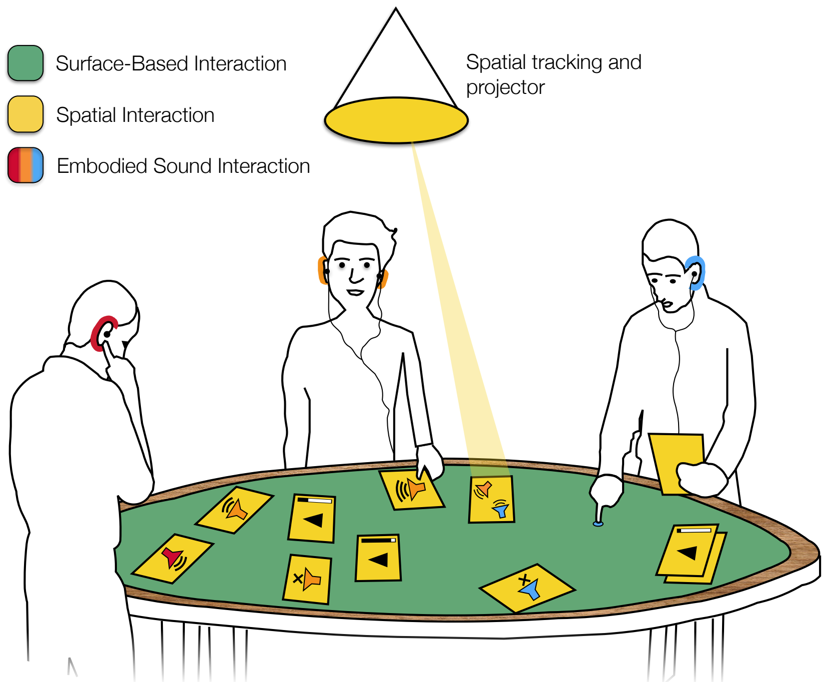


Figure 1.1: The three research directions of this thesis each contributes novel devices and interaction concepts for interaction with visual and sound content. All combined in a meeting scenario during a fluid collaboration.

Surface-Based Interaction (Green): **This thesis contributes a device concept and an integrated set of interaction techniques that support fluid collaboration on the very same tabletop surface. This novel tabletop surface supports each user with personalized input and output.**

Spatial Interaction (Yellow): **Spatially aware paper-like displays allow a group of people to interact with the digital content (e.g., sound or videos) by physically moving paper-like displays in space, similar how to we use paper documents.**

Embodied Sound Interaction (Red, Orange, and Blue): **Our concept, called EarPut, augments behind-the-ear-worn devices, such as earphones, for private sound interaction. Multiple paper-like displays allow spatial interaction with multiple videos or sound sources privately and collaboratively. Both concepts allow for direct manipulation of sound sources.**

et al., 1985] of visual content. There is a lack of suitable devices and interaction concepts for direct manipulation of sound content. Sound is either heard individually (e.g., earphones) or collaboratively (e.g., speakers). Switching between individual and collaborative sound simultaneously is, however, only limited possible, and this hinders a collaboration with sound.

With this focus on collaborative sound interaction, we contribute novel devices and interaction concepts for direct manipulation of multiple sound sources. Particularly with the focus on fluid collaboration, our goal is to allow each user to manipulate and control multiple sound sources in an individual and in a collaborative way. Our contribution proposes to control sound individually with *body-based interaction*, by touching the area where sound is perceived such as one's own ear. Another way of interacting with sound both individually and collaboratively is *spatial interaction*. This is done by spatially moving paper-like displays in 3-D space. Both interaction styles are *embodied* [Dourish, 2004], because of their body-centric and physical interaction. Both embodied interaction styles are needed to support fluid sound collaboration that expects a fluent switch between the individual and the collaborative way of interacting with sound.

In the following, we elaborate which type of collaboration our novel devices and interaction concepts are supporting. Then we briefly summarize the contribution of this thesis by visiting all three research directions. Next, we present our research methodology and conclude this chapter with our publications.

1.2 Background and Research Context

This dissertation focuses on the general research area of human-computer interaction (HCI). HCI studies the design of interactions between humans and computers. HCI has various research subareas; one of them focuses on concepts, methods, and corresponding tools for computer-supported cooperative work (see Definition 1).

Definition 1 (Computer-supported cooperative work (CSCW))
Computer-supported cooperative work is computer-assisted coordinated activity carried out by groups of collaborating individuals. [Baecker *et al.*, 1995, p.141]

The research context of this thesis is depicted in Figure 1.2. Within CSCW, this thesis focuses on a particular type of collaboration called *mixed-focus collaboration*, which will be explained in the following in detail. Collaboration in general is dependent on space and time [Dix *et al.*, 1998, p.465][Johansen, 1988]. This thesis contributes collaborative

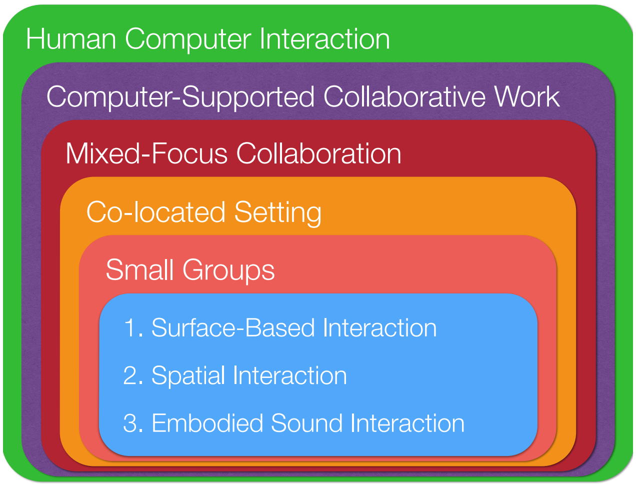


Figure 1.2: Research context of this thesis

systems that focus on face-to-face interactions in the same time and location (see Figure 1.3). Thus, the work focuses on co-located groups with a small size (e.g., two to six users). Thereby working together in close phase social distance (1.2 to 2.1 m) [Hall and Hall, 1969] with an interactive system.

This small group of people (e.g., two to six users) is working in a specific type of collaboration, namely, mixed-focus collaboration, which often occurs during group work. This type of collaboration involves both individual and shared activities. While working in a group, users usually require a shared common view to have the same reference point for discussion. While working individually, users partition and occupy their own respective spaces. Even when working individually, people try to maintain an understanding of the activities of other collaborators. In these mixed-focus situations, users have to constantly switch between individual work and group work. This collaboration type has been first introduced and named by Gutwin in 1998:

Definition 2 (Mixed-focus collaboration)
A third kind of collaboration flips between same-view and different-view situations. We call this kind of interaction mixed-focus collaboration: individual and shared activities within the workspace are interleaved, and learners periodically shift their attention back and forth between separate and shared views of the workspace. [Gutwin *et al.*, 1995, p. 6]

During mixed-focus collaboration, different coupling (see Definition 3) styles exist that have been introduced by Tang *et al.* [2006]. The spectrum of coupling styles ranges from *tight coupling*, when all collaborators are actively working together on the same problem, to *loose coupling*, when collaborators are independently working on separate problems.

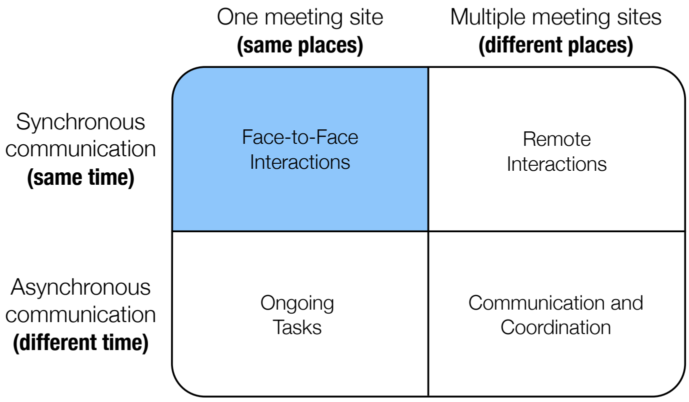


Figure 1.3: A time/space matrix is used to classify groupware systems. This thesis contributes collaborative systems that focus on face-to-face interactions, being in the same time and also in the same space.

Many more *mixed coupling* styles exist in between these two ends [Isenberg *et al.*, 2012]. Different coupling styles require different support in visualization and interaction. While discussing mixed-focus collaboration, we specifically refer to collaboration in which users are working in close physical proximity [Kiesler and Cummings, 2002; Xiao, 2005].

Definition 3 (Coupling)
Coupling is the amount of work that one person can do before they require discussion, instruction, action, information, or consultation with another person. [Gutwin and Greenberg, 2002, p. 20]

Different coupling types during collaboration have different inherent limitations and shortcomings. Group work usually requires workspace coordination (see Definition 4) and awareness (see Definition 5) of group activities, especially in mixed-focus collaboration.

Definition 4 (Workspace coordination)
Workspace coordination is the management of access to and transfer of shared resources. [Schmidt and Simonee, 1996; Tang *et al.*, 2006]

Definition 5 (Awareness)
Awareness is an understanding of the activities of others, which provides a context for your own activity. [Dourish and Bellotti, 1992, p. 107]

Insufficient support of workspace coordination on one interactive surface frequently results in workspace interference (see Definition 6). One example is access conflicts on a shared surface, when access to a particular interface element is disputed [Morris *et al.*, 2006]. However, these conflicts require collaborators to coordinate their interactions through, for example, partitioning the surface into dedicated personal and group territories [Scott *et al.*, 2004; Sugimoto *et al.*, 2004]. Although this partitioning alleviates interference, it constrains each user in both *interaction* and *screen space*.

Definition 6 (Workspace interference)
Workspace interference is the act of one person hindering, obstructing, or impeding another's view or actions on a single shared display. [Zanella and Greenberg, 2001]

This thesis focuses on allowing users to fluently switch between individual and group work during fluid collaboration (see Definition 7). Thereby, novel devices and interaction concepts are required to support fluid collaboration.

Definition 7 (Fluid collaboration)
Fluid collaboration is a computer-supported cooperative work where a small co-located group of people (two-to-six) is involved in both individual and group work, with the ability to fluently (seamlessly and intuitively) switch between these.

In the following, we present the contributions of this thesis.

1.3 Contribution and Thesis Structure

This thesis presents three different research directions that all contribute to fluid collaboration (see Figure 1.4). First, *surface-based interaction* focuses on a fluid collaboration around an interactive multi-view tabletop, allowing personalized input and output for each user. Second, *spatial interaction* transfers well-known affordances from paper to the digital world by combining video navigation with multiple paper-like displays, thereby allowing tangible interaction and spatial structuring (e.g., moving and piling) and playback of multiple videos in physical space. Last, *embodied sound interaction* proposes novel ways of directly manipulating sound individually and collaboratively by the use of body-based and spatial interactions. In the following, these three research directions are presented one by one.

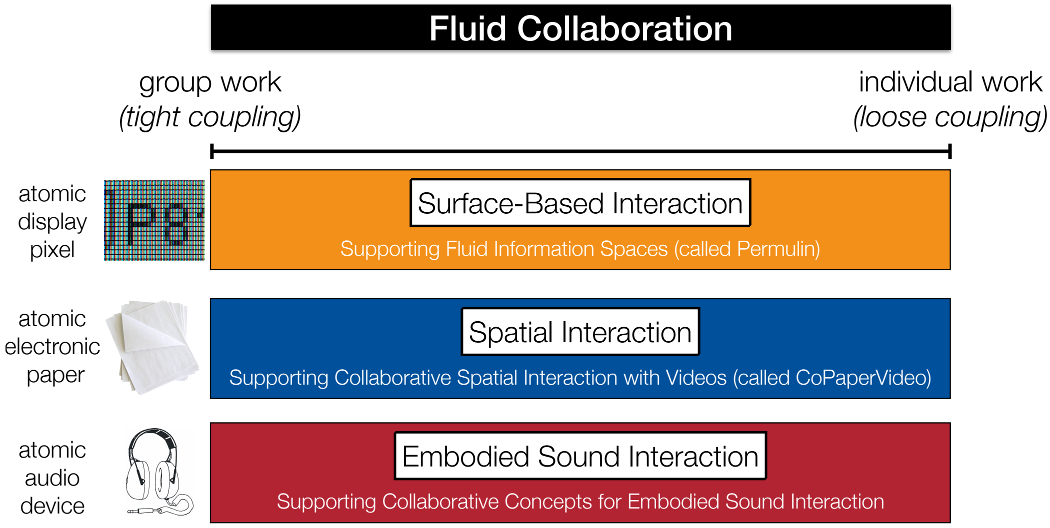


Figure 1.4: Overview of the research direction where novel devices and interaction concepts for fluid collaboration are contributed. The three areas are also reflected in the structure of this thesis.

1.3.1 Surface-Based Interaction

Group work exists in nearly all working areas. Thereby, users are mostly seated around a table to work together in order to make decisions. Nowadays, supportive digital information can be displayed while sitting around a digital interactive table, called *tabletop*, with several collaborators. Soon all meeting rooms will feature such tabletops, and it will be a promising starting point for meetings and collaborations. However, these tables still have their limitations: (1) collaborators find it hard to switch between individual work and group work, because on an always *shared screen and with constant awareness*, individual space can hardly be created, and this leads to (2) a *limited interaction space and screen clutter*, which mostly generate (3) *workspace interference and access conflicts* during interaction. All these problems exist while collaboratively working with horizontal surfaces. Hereby, novel devices and interaction concepts need to be developed in order to improve these limitations.

In order to *support individual and group work on the very same interactive horizontal surface, each user needs a private or a group view*. For group work, users around the surface should have a common group view in order to work together. In addition, this surface should provide private views for individual work to each user. Private views can cover only a small part of the display, or they can cover the full screen or even consist only of private elements. In addition, any private element or private view can be shared to the common group view.

To implement this concept, it requires a change of our understanding of the atomic display pixel. Currently, each display pixel shows content that is visible to all users and all users can interact with it. In order to support private views, private elements, and a group view simultaneously, *each pixel should support different levels of visualization*– ranging from private visualization of private views or elements to each user, as well as a shared common visualization to all users, which in the following will be called *personalized or group output.*

In addition to the output, input is of vital importance. If output is provided privately, each pixel needs to know who is interacting with it, so that content visualized privately can only be accessed and manipulated by the corresponding user. Based on this assumption, *each display pixel has to support both (1) personalized output and (2) personalized input.*

In chapter 2, we present Permulin, an interactive multi-view tabletop that allows users to fluently switch between individual and group work (see Figure 1.5). We further present interaction techniques that support a fuller spectrum between different coupling styles that appear during co-located collaborations, such as tight and loose coupling.

Support of fluid collaboration on a single display, such as an interactive tabletop, is the focus of surface-based interaction. Concepts discussing support of fluid collaboration with videos on multiple paper-like displays, called spatial interaction, is presented in the following.

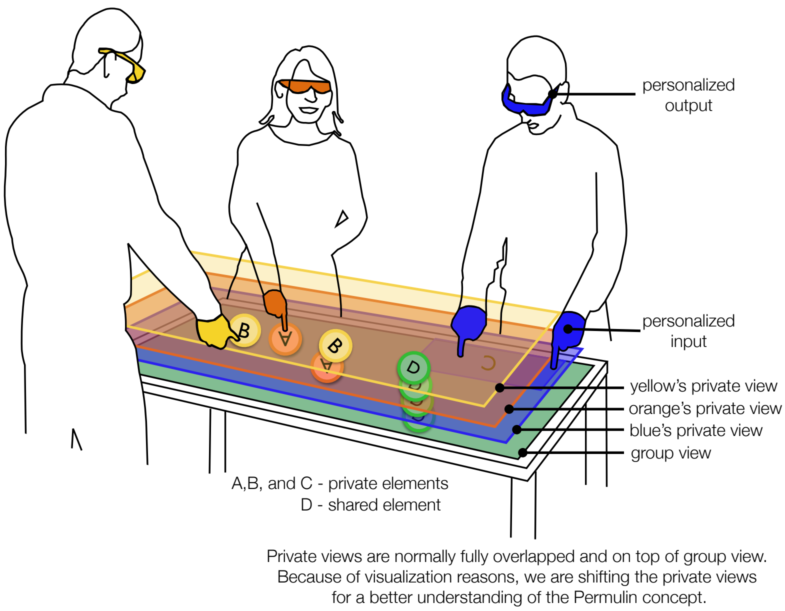


Figure 1.5: Permulin: an integrated set of interaction and visualization techniques for multi-view tabletops to support co-located collaboration across a wide variety of collaborative coupling styles

1.3.2 Spatial Interaction

People from a variety of professional backgrounds are confronted daily with large amounts of video footage that they must sift through and make sense of. TV news editors, for example, have to deal with approximately 30 hours of video material offered per news agency per day (e.g., Reuters). The YouTube era extends these tasks of sifting and making sense out of many videos to the general population, for hobby and scholarly activities. These examples show that *active video work* with large amounts of video material (as opposed to passive watching of a single video) is a daily routine for many people.

Sifting and sense-making of paper-based information is a well-researched field. Knowledge gathered in this research field can be transferred to active video work. Research shows that the key is using not only one but multiple documents or sheets of paper simultaneously, in order to *manipulate and organize information in physical space*. Furthermore, it has been proven to effectively support comparison, overview generation, and better orientation, particularly because of the use of spatial interactions [Kirsh, 1995; Sellen and Harper, 2001]. All these activities are also of paramount importance when working with videos.

Compared to paper-based practices, today's user interfaces for collaborative active video work still have their shortcomings:

(1) Multiple users are mostly restricted to a single screen, so collaborative video browsing is limited in a co-located scenario.

(2) While standard navigation techniques for videos (e.g., play, pause, stop, and seeking on a laptop or mobile device) have their obvious benefits, they lack the effectiveness of physical interaction [Kirsh, 1995; Mackay and Pagani, 1994; Sellen and Harper, 2001] for spatially structuring videos.

(3) The traditional "one video at a time" paradigm, where only one video at a time is playing, does not leverage the whole spectrum of human perception. While humans are able to focus only on limited information, they are able to grasp a much higher amount of information in the periphery, which is helpful in getting an overview and structuring.

We assume that these affordances of working with multiple documents in physical space can be effectively transferred to the domain of video. We advocate a paradigm for videos that consists of using multiple videos simultaneously (see Figure 1.6), similar to how we lay out multiple printed documents on our desk. Hereby, we investigate how interactions known from physical documents can be transferred to the world of videos and fitted to emerging mobile computing devices as lightweight, thin, and flexible as paper, further

[2] http://www.reuters.com/

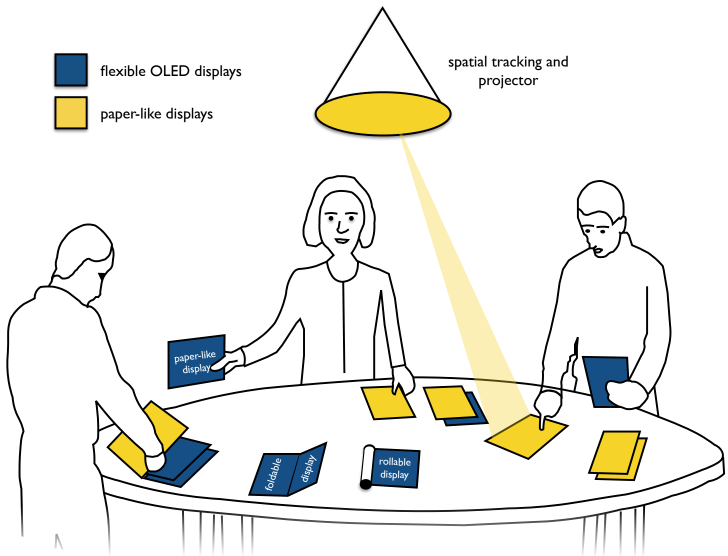


Figure 1.6: Spatial interaction with spatially aware displays. (Yellow) Lightweight paper-like displays that are tracked in space and content is projected onto the display. (Blue) Flexible OLED displays that feature high resolution and are also spatially aware. Both technologies are used together.

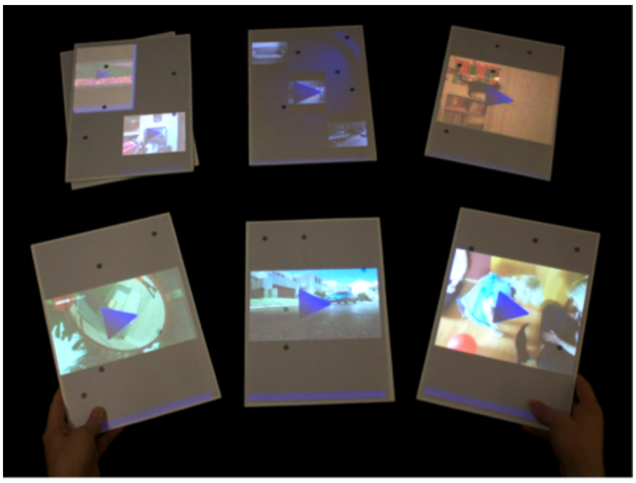


Figure 1.7: CoPaperVideo: a system for collaboratively interacting with multiple videos in physical space

called paper-like displays [Co and Pashenkov, 2008; Crawford, 2005]. These paper-like displays feature a *high display refresh rate* that supports visualization of high dynamic digital content (e.g., videos).

In this spatial interaction part, we present CoPaperVideo, a coherent system that allows multiple users to play back and navigate through videos and collections of videos with multiple physical displays (see Figure 1.7). Furthermore, we introduce spatial interactions that enable users to create an overview of multiple videos, structure and organize video content, and manage this content on multiple displays.

1.3.3 Embodied Sound Interaction

In the previous two research directions, we explored new ways of interacting with visual output. Visual interfaces are effective during collaboration particularly because they allow for direct manipulation [Hutchins *et al.*, 1985] of the represented content. In terms of sound interaction, direct manipulation of sound, is to our knowledge, only limitedly possible because of lack of suitable devices and interaction concepts. In this chapter of the thesis, we focus on direct manipulative sound interaction by investigating novel devices and interaction concepts for sound. Particularly with the focus of fluid collaboration, the goal is to allow each user to manipulate and control multiple sound sources in both individual and collaborative way.

Individual sound is perceived through our ears. For direct manipulation with perceived sound, we propose a novel device concept, called EarPut. EarPut is a smart earpiece that augments accessories that are placed or worn behind the ear (e.g., earphones). It allows users to instrument their ear, as an interactive surface to enable eyes-free, mobile interaction with the ear. This allows each user to *directly control their individual sound by touching their own ear* suitable for individual interaction with sound. Since EarPut primarily focuses on input, we envision it as a companion device that piggybacks onto existing feedback mechanisms, for example, to wirelessly trigger auditory or vibrotactile feedback through actuators of a smart phone.

In order to provide each user the ability to *focus and control multiple sound sources individually and collaboratively* as well as fluently switching between both collaboration styles, we contribute spatial interaction with sound. Thereby, sound is virtually placed in 3-D space and is physically associated with spatially aware paper-like displays that embed videos. This allows each user to control the volume by physically moving paper-like displays in space. The proposed spatial interaction with sound allows each user to control sound in a direct way as how we are used to while structuring and organizing paper documents.

These two novel devices and interaction concepts combined, advocate a novel collaboration paradigm for directly manipulating multiple sound sources similar with that we used to do during interacting with visual content. Thereby, the presented contributions propose to control sound individually with either *body-based interaction* by touching one's own ear or with *spatial interaction* by spatially moving paper-like displays in 3D space. Both interaction styles combined are named **embodied** [Dourish, 2004] because of their body-centric and physical interaction. Both embodied interaction styles are combined to

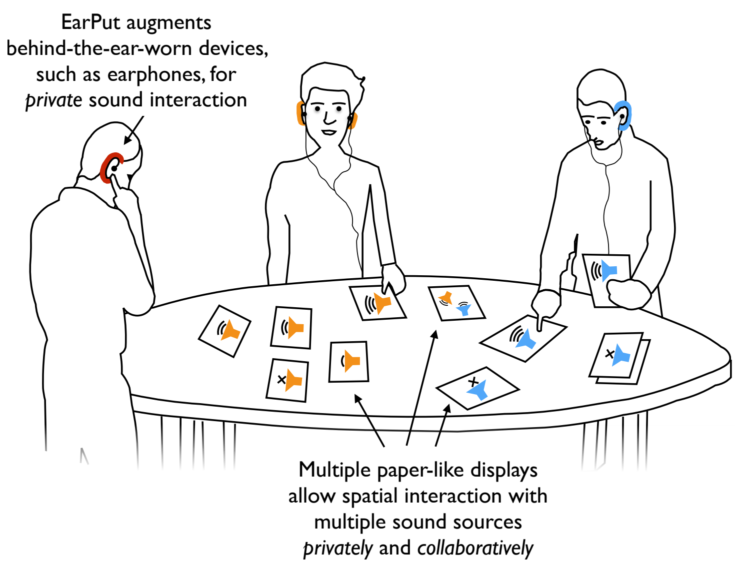


Figure 1.8: The concept of embodied sound interaction. Users can touch their own ear to manipulate sound privately. Paper-like displays allow a group of people to spatially control parallel sound sources.

support a fluent switch between individually and collaboratively interacting with sound (see Figure 1.8). Users are wearing earphones in order to have virtual 3-D sound, which is needed for allowing spatial interaction. These earphones can unobstrusively be augmented with EarPut to allow each user a private direct sound control.

The contributions regarding *embodied sound interaction* will be presented in the chapter 4. Although our main scenarios feature sound from videos, our concepts are also suitable for situations with sound-only output.

1.4 Research Methodology

This thesis research was conducted in the following directions as discussed above: (1) surface-based interaction, (2) spatial interaction, and (3) embodied sound interaction. Throughout these research directions, empirical methods were employed for understanding users as well as real-world challenges and to therefore develop suitable, supportive, easy-to-use and enjoyable interaction concepts for novel devices for individual and group work (see Figure 1.9 for a schematic overview).

In all three research directions, results of qualitative explorations with users informed our hypothesis and requirements and provided initial insights into the proposed novel devices. During this iterative design of novel devices and interaction concepts, a user-centered design approach was used [Norman and Draper, 1986].

All predefined research questions and a hypothesis were explored and verified using qualitative and quantitative methods, thereby allowing validation and verification of the presented contributions of this thesis. Qualitative methodology in the form of an explorative approach has been used to allow for accessing behavior, thoughts, mental models, as well as feelings of users. In particular, research directions such as surface-based interaction and embodied sound interaction used quantitative methods for the evaluation of novel devices with previously defined hypothesis. Furthermore, quantitative methods allowed for measuring precision and effectiveness of users during interaction.

1.5 Publications

Parts of this thesis are partially published in proceedings of international peer-reviewed conferences, such as the ACM SIGCHI Conference on Human Factors in Computing Systems (*CHI*) and ACM Multimedia (*MM*).

The chapter "Surface-Based Interaction" is partially published in Lissermann *et al.* [2013b,c, 2014b]. For the chapter "Spatial Interaction," parts except the sound interaction techniques have partially been published in Lissermann *et al.* [2012a,b]. The "Embodied Sound Interaction" chapter presents collaborative sound interaction concepts and has partially been published in two separate publications. The first publication focuses on individual interaction with sound by presenting a device called EarPut [Lissermann

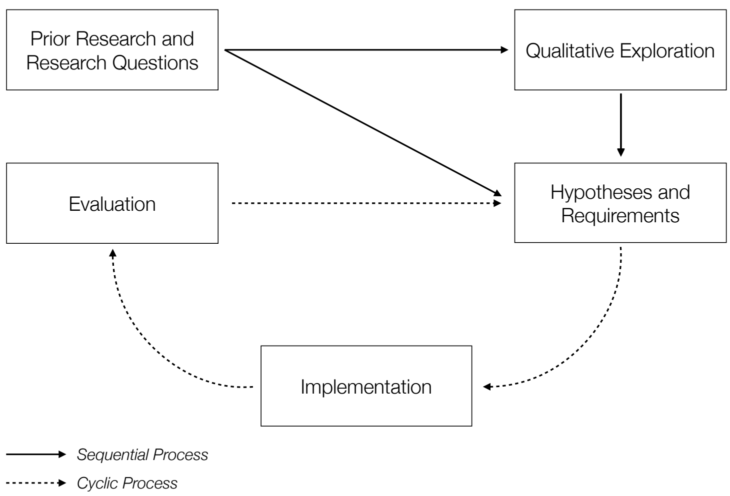


Figure 1.9: Overview of the research methodology

et al., 2013a, 2014a]. The second publication presents spatial interaction concepts with parallel sound sources [Lissermann *et al.*, 2012a].

In addition, I have contributed as the second author to interaction concepts for resizable displays [Steimle *et al.*, 2012], such as rollable displays [Khalilbeigi *et al.*, 2010, 2011] or foldable displays [Khalilbeigi *et al.*, 2012]. However, these publications are of lesser focus in this thesis.

2

Surface-Based Interaction

Contents

Nowadays, nearly every *co-located collaboration* (e.g., meetings) needs digital information. However, current available devices such as mobile devices or shared display do not provide the needed visibility and access of this information to all collaborators.

Interactive tables can provide a solution to these limitations. Interactive tables (also called *tabletops*) have been the focus of researchers since decades. Tabletops allow collaborators to sit around this table similar as they are used to do around a normal table. They also allow all collaborators to interact with multi-touch input on a horizontal interactive screen. Similar to a normal table, all users have a common shared view, and all collaborators can access content shown on the tabletop.

Tabletops, however, still have unsolved challenges: Similar as working with multiple physical documents, users work with multiple digital windows. This requires spatial partitioning of the screen into several smaller views [Isenberg *et al.*, 2012]. Especially when working in fluid collaboration (see Definition 7), spreading these views can lead to either *interference* or *limited space*. Another drawback of an always shared view of a tabletop is *distraction*. While working individually, users might want to focus on their activity without seeing what other collaborators are doing.

In order to overcome these challenges and to support fluid collaboration on an interactive surface, we contribute a novel device and interaction concept we call *Permulin*. Permulin is a tabletop that allows *personalized output*; each user can switch between personalized and shared full screen view. In addition, Permulin provides *personalized input*, where each touch point is correspondingly mapped to the user. Personalized in- and output combined allow interaction techniques for Permulin, which support a fluent switch between individual and group work on an interactive surface.

This chapter has partially been published at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) Lissermann *et al.* [2014b] and Lissermann *et al.* [2013b,c]. The remainder of this chapter is structured as follows: First in section 2.1 we describe the scope of this chapter. In section 2.2, we present related works and a design space of surface-based interaction. Based on requirements for personalized input and output presented in the related work, we introduce the *Permulin* concept, followed by an explanation of technical realization and application scenarios of our system, which are described in section 2.3. In section 2.4, novel interaction and visualization techniques for fluid collaboration on interactive tabletops are described. Results from two user-centric evaluation studies are reported in section 2.5. They show that Permulin supports fluent transitions between individual and group work and exhibits unique awareness properties that allow participants to be highly aware of each other during tightly coupled collaboration while being able to unobtrusively perform individual work during loosely coupled collaboration.

In summary the main contributions of this chapter are as follows:

1. First, a profound review of the related work in the area of an interactive surface with personalized input and output are presented. Based on the related work, we contribute a fluid collaboration design space for interactive surfaces.

2. Second, design and development of *Permulin*, which is an integrated set of novel interaction and visualization techniques for fluid collaboration on a tabletops with personalized input and output are described.

3. Last, two user studies that focus on collaboration on a tabletops with personalized input and output are presented. Our results show that Permulin allows users to transition fluently between individual and group work on an interactive surface.

2.1 Scope of Surface-Based Interaction

The scope of this chapter is interactive surfaces in specific interactive tabletops. While focusing on *surface-based interaction* in the form of multi-touch input, we provide novel interaction and visualization techniques for such interactive surfaces.

In co-located collaboration on surfaces such as digital tabletop, collaborators usually interact on one common shared view. Similar as working with multiple physical documents, users are working with multiple digital windows. This requires spatial partitioning of the screen into several smaller views [Isenberg *et al.*, 2012]. Thus, the use of shared common view on a tabletop, is likely to lead to either interference or limited space when users work in mixed-focus collaboration (see Definition 2).

Tabletops support only a shared view, allowing only a single representation of their full screen. In contrast, *multi-view* hardware, which can show two or more different images at the same spatial location [Harrison and Hudson, 2011; Kim *et al.*, 2012; Matusik *et al.*, 2008; Mistry, 2009], could allow rendering of personal views for different users at the same location on the very same screen.

Definition 8 (Multi-view tabletop)
Multi-view tabletop is a tabletop that provides both personalized in- and output to each user. Personalized output allow each user to switch between personalized or shared visualized small or full-screen view. During personalized input, each touch point is correspondingly mapped to the user.

For this reason, multi-view tabletop (see Definition 8) hardware seems highly promising to support fluent transitions between different coupling styles [Tang *et al.*, 2006] in mixed-focus collaboration. The spectrum of coupling styles ranges from *tight coupling*, when all collaborators are actively working together on the same problem, to *loose coupling*, when collaborators are independently working on separate problems. Many more mixed coupling styles exist in-between these two ends [Isenberg *et al.*, 2012]. In this chapter, we show that different coupling styles requires different visibility and access that lead to different visualization and interaction support.

Pioneering research has introduced first, promising principles for multi-view tabletop interfaces, which overlay additional private information on a shared view [Agrawala *et al.*, 1997; Karnik *et al.*, 2012; Matsushita *et al.*, 2004; Shoemaker and Inkpen, 2001]. Compared with classic tabletop interfaces, this provided additional personalized support during tightly coupled collaboration.

We aim to provide support for a considerably fuller spectrum of collaborative coupling styles with a novel concept of multi-view tabletops we call *Permulin*, covering both ends of the spectrum. Moreover, Permulin supports fluent and seamless transitions between these styles. Before presenting our visualization and interaction techniques, we first visit related work.

2.2 Related Work

For a systematic review of the related work, we first present the design space for fluid collaboration on interactive surfaces (see Figure 2.1), thereby mainly focusing on related work that have provided support for co-located collaboration.

The design space is a composition of input and output on a tabletop screen. Different tabletops exist that feature either a private or a public input. For *public input*, each touch point is anonymous for the system, whereas in *private input*, each touch point is correspondingly mapped to the user. A similar approach has been followed for output. *Public output* allows only a single output on the screen, whereas a *private output* allows each user to have a private (further called personalized) output on the tabletop screen.

This results in four different areas that allowed to systematically structure previous related work and show our contributions:

Shared Information Space: We present research projects that feature public input and output while providing a shared information space for collaboration. These works will be discussed in subsection 2.2.1.

Personalized Input: Interactive tabletops that map the input to a user are called tabletops with *personalized input*. They are presented in subsection 2.2.2.

Personalized Output: Previous systems are presented that provide each user with a personalized output. These systems are further called tabletops with *personalized output* and are presented in more detail in subsection 2.2.3.

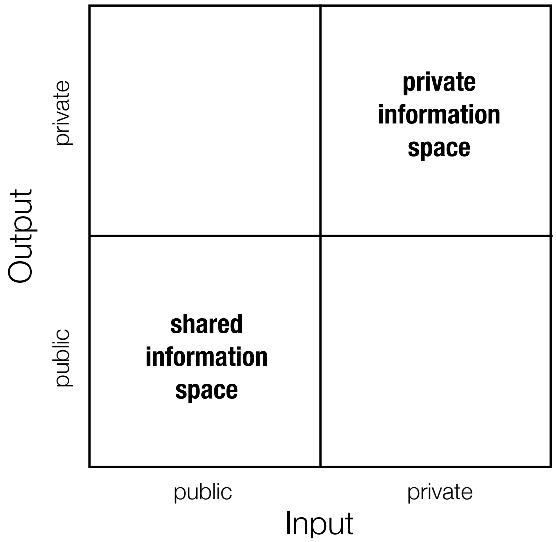


Figure 2.1: Design space for surface-based interaction

Private information space: Private information space provides both *personalized input and output* on a tabletop. Allowing each user to have personalized views only visible to them and private interaction on these views. These systems are presented in subsection 2.2.4.

2.2.1 Shared Information Space

Information on the screen is visible and accessible to all collaborators and is, due to that, shared. We visit research projects that focused on interactive surfaces with public input and output. These devices also support collaboration. They, however, have several limitations that are described in the following paragraphs.

Interactive tabletops have been investigated by researchers since the early 1990s. Digital Desk [Wellner, 1991] was one of the first systems introduced for individual users that allowed multi-touch interaction on a projected surface. Since then, various hardware solutions have been introduced [Müller-Tomfelde, 2010] that are also suitable for collaboration in a group of people and have been investigated in group settings during collaboration [Scott, 2005].

Group work on interactive surfaces, while being in close physical proximity [Kiesler and Cummings, 2002; Xiao, 2005], usually requires coordination of group activities, especially in mixed-focus collaboration. Insufficient support of workspace coordination (see Definition 4) on one interactive surfaces frequently results in interference (see Definition 6). One example is access conflicts on a shared surface, when access to a particular interface element is disputed [Morris *et al.*, 2006]. However, these conflicts require collaborators to coordinate their interactions through, for example partitioning the surface into dedicated personal and group territories [Scott *et al.*, 2004; Tse *et al.*, 2004]. Although this partitioning alleviates interference, it constrains each user in both *interaction* and *screen space.*

Definition 9 (Multi-Device Environment (MDE))
Within an MDE, the term device is used to refer to a laptop, tablet, large display, etc., each driven by an independent, but networked system. [Biehl and Bailey, 2006]

Additional devices (e.g., laptop, tablets, or large displays) are another approach to overcome limited space and interference in mixed-focus collaboration. This is also known as multi-device environment (see Definition 9).

WeSpace [Wigdor *et al.*, 2009], Caretta [Sugimoto *et al.*, 2004], or MobiSurf [Seifert *et al.*, 2012] are good examples that combine interactive surfaces for group work with personal devices for private interaction. Some have focused on providing interaction techniques to control and manage applications in such environments [Biehl and Bailey, 2006; Rekimoto and Saitoh, 1999]. Others have investigated teamwork in single- and multi-device environments [Wallace *et al.*, 2009].

However, all these approaches require the collaborators to *switch their attention* between the surface and the secondary device [Cauchard *et al.*, 2011; Su and Bailey, 2005; Tan and Czerwinski, 2003]. Furthermore, increasing the size or number of displays is not necessarily an advantage [Ryall *et al.*, 2004].

2.2.2 Personalized Input

Interactive conventional tabletops can recognize multiple touch input. Touch input, however, are anonymous among users so that all users, input is treated equally. For example, in a collaborative painting task, if one user selects a color, it forces all users around the table to use this color. Tabletops with personalized input support a direct one-to-one mapping from touch input to the user, allowing the system to know which user is interacting with it. This personalized input can support another way of collaboration and can lead to a better capability to collaborate.

Various techniques have been used to provide a direct mapping from a touch input to a corresponding user. For example, different sensors have been used such as capacitive sensors, proximity sensors, or pressure sensors to provide personalized input. Other approaches for personalized input are: additional devices such as additional camera mounted above an interactive table, additional accessories such as mobile phones, or specific gloves. Figure 2.2 shows an overview of the various techniques used to provide personalized input. In subsubsection 2.2.2.1, requirements are presented for personalized input. The different approaches are then more closely presented and evaluated with respect to the introduced requirements.

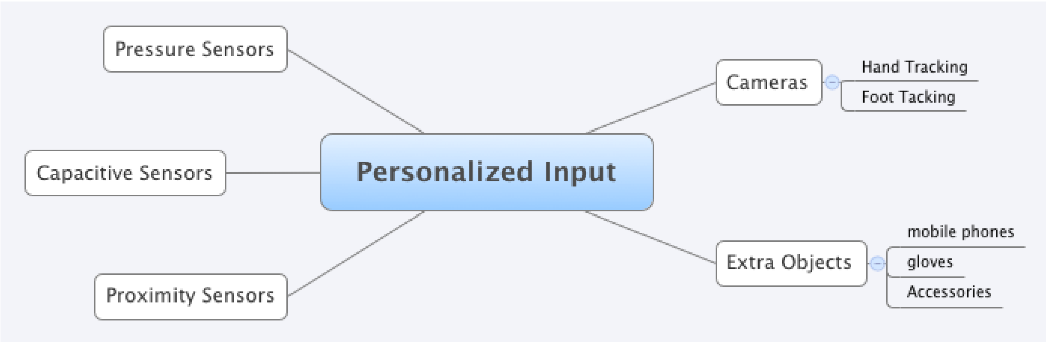


Figure 2.2: Overview of the techniques previously used to provide personalized input.

2.2.2.1 Requirements

In order to compare previous approaches to realize personalized input, we present previously introduced requirements for touch-based personalized input (see I1-I7). These requirements have been previously introduced by: Dietz and Leigh [2001] (presented requirement I1, I2, I3 and I4), Harrison *et al.* [2012b] (presented requirement I2 and I5), Ramakers *et al.* [2012] (presented requirement I2, I4, I6, I7), and Schmidt *et al.* [2010] (presented requirement I4, I6).

I1: Simultaneous Input Multiple touch points need to be processed, recognized, and mapped to a user simultaneously. This *user detection* allows a user a direct and natural interaction.

I2: Multiple Users The system should support a high amount of simultaneous users with personalized input functionality. However, the number must be distinguished between (1) *user profiles* [1] and (2) *user identification.* [2]

For example, if a system recognizes 10 users via previously stored user profiles but due to other system properties only two users can be identified during interaction, we would only consider the number two for better comparison.

[1] The number of users stored in the system.

[2] The number of user that can be clearly recognized and mapped to a specific user during interaction. User identification allows the system to authenticate a user without users input e.g. password.

I3: Robustness The user detection and user identification should be robust. Physical objects put on the surface should not interfere with normal operation. The system should withstand normal use without frequent repair or re-calibration.

We extend the previously mentioned factors and requirement a system to stay functional in different light conditions (dimmed or room light), and therefore extends the term of *robustness*.

I4: No User Instrumentation A system should avoid as far as possible, that users need to be instrumented and carry additional equipment to allow personalized input (e.g., a transmitter or special pens). Specific gestures, for example, password input that restrict the user in starting to interact with the device or that change the working process of the user, are also considered as user instrumentation.

I5: Easy Deployment Mobility of the system is an important requirement while dealing with personalized input. We consider a system mobile when the complete system is easy to deploy in another environment. Hereby, the system should have a certain compactness and a minimal instrumentation of the environment and low energy consumption. This is especially interesting in a very adaptive office environment.

I6: Sustainable User Association Sustainable user association refers to the ability of a system to maintain users' identification even when they leave and reenter the sensor area.

I7: Free User Positioning The system should support a free positioning of the user around the tabletop and thereby especially a flexible viewing angle.

In the following, we will revisit previous research approaches for personalized input with respect to the stated requirements.

2.2.2.2 Capacitive Sensors

Capacitive techniques are based on the principle of capacitive coupling. Hereby, energy load from one object is transferred to another object which leads to a voltage drop that can be measured. These techniques rely on the fact that human skin is conductive to a certain extent.

In essence, the following principle is applied: A voltage is applied to a conductive object. A touch of a human finger leads to a change in voltage because the finger hereby is an electrical conductor. This voltage change can be measured. This knowledge can be used to recognize position and identification of the user. In the following, different previous works are discussed that are related to capacitive techniques.

Diamond Touch

DiamondTouch presented by Dietz and Leigh [2001] was the first commercially available tabletop that assigns multiple touch points to the user. DiamondTouch has a grid pattern of tiny antennas that are isolated, and each transmits a uniquely identifiable signal. In addition, each chair of a user is connected to the system via a cable. The two-dimensional antenna pattern in the form of a diamond has been developed to identify touch position (see Figure 2.3). Through this pattern and a clear signal that is sent out from each antenna, a recognition of multiple touch points can be generated. By touching the table, the circuit of table through the user, the chair to the system is closed so that the system is also able to recognize who touched the table.

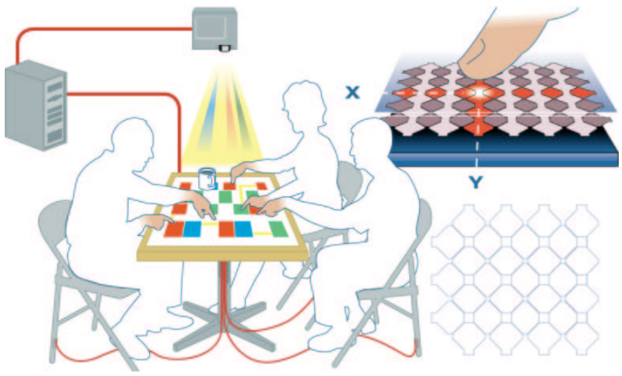


Figure 2.3: Overview of the DiamondTouch system. Touch position is calculated though a grid antenna pattern. User association works though connected chairs with the system and a closing circuit through the human body while touching the table. (Figure copied from `http://resenv.media.mit.edu/classes/MAS965/readings/DiamondTouch.pdf`)

Evaluation: The DiamondTouch is the first technology that detects multiple touch points (I1 fulfilled, in the following abbreviated with I1). The number of the simultaneous users, however, is dependent to the number of chairs (I2), which needs to be connected to the system. An additional user that wants to interact with the system cannot easily be recognized. DiamondTouch dependents on chairs that hinders the mobility and deployment of the system (I5 not fulfilled). This results in a lack of a compactness and easy deployment. Dietz and Leigh, however, emphasize the low production costs. Through the use of chairs, a unique user mapping exists, if the user switches their chair, also map to a user switch because users are identified via chairs (I6 partially fulfilled). In addition, the fixed sitting position of the user do not allow free choice of position (I7 not fulfilled). A positive note is the low instrumentation of the user (I4) and that the system is robust and does not need to be repaired or recalibrated (I3).

Figure 2.4: The figure shows the capacity profiles for different grasping gestures. This figure was taken from Sato *et al.* [2012].

Capacitive Fingerprinting

Capacitive fingerprinting was first introduced by Harrison *et al.* [2012b]. It uses the electrical properties of the human body to differentiate between different users. Each human body differs, for example, in bone density, muscle mass, or blood volume. Outside the body, users differ in their choice of clothing or shoes. By varying these properties, different user profiles can be created. This can be used to identify the user with his capacitive fingerprint.

The basis used for user identification is called Swept Frequency Capacitive Sensing (SFCS) [Sato *et al.*, 2012] , which was also developed by Sato, Pupyrev, and Harrison. Conventional capacitive touch technology measures typically only whether a touch has occurred or not by measuring electrical signal at a fixed frequency. SFCS extended this procedure by measuring "the response to capacitive human touch in a range of frequencies." Objects connected to an electrical signal respond differently when touched by the user at different frequencies. Thus, different gestures such as touch, grasp with two fingers, or hold with all fingers can simply be recognized with a single electrode that is connected to the object.

With a support vector machine (SVM) capacitive profiles for different gestures on the objects can be learned and distinguished (see Figure 2.4), for example, for a door knob, various single- and multi-finger gestures can be distinguished. In order to identify a respective gesture, previously trained capacity profiles for the different gestures need to be prerecorded and stored in the system. The SVM is capable of mapping the current performing gesture by the user to the previously recorded gestures.

In the context of the capacitive fingerprint, SFCS is used not only to detect different gestures but also to differentiate different users touching a surface. Hereby, an existing touch display is extended with a special film, which allows the measurement of capacitive profiles while the user is touching the system without restricting the user in the use of the touch display (see Figure 2.5). A user touches the film via SFCS, a capacity profile depending on the frequency can be determined . A user profile can be calculated via a SVM compared with previously stored capacity profiles of all users. This is successfully identified when the measured capacity profile is sufficiently overlapping with the ones stored previously.

Evaluation: In the current state of development , the technique of Harrison, Sato, and Poupyrev does not allow simultaneous personalization of multiple touch points because the system consists of only one electrode (I1 not fulfilled). However, users can be differentiated robustly (I3) and are identified to a certain extent even (I2) without instrumentation (I4). In addition, users are assigned even if they leave the sensor region because the user identification is based on the user specific capacity profile (I6). Thus, the users can freely choose their position relative to the tabletop system (I7). Furthermore, the capacitive fingerprinting approach is also easily employable, as existing displays need to be extended only by a thin layer of electrodes (I5) . According to Harrison, Sato, and Poupyrev, SFCS interferes with capacitive touch technology.

2.2.2.3 Proximity Sensors

The techniques outlined below are based on the use of multiple proximity sensors. A proximity sensor can determine the presence of an object in a contact-free manner. A simple example of a proximity sensor is light barriers, which can detect interruption of a light beam.

Most of the systems presented in the following are equipped with sensors that are built in the frame of the table and assume that activation of the sensor is equivalent to the presence of a user. Activation of multiple sensors can therefore be used to track whether a user is approaching or moving around the table.

Tănase et al. and Walther-Franks et al.

Tanase *et al.* [2008] presented a system that used twelve proximity sensors mounted onto the frame of a tabletop (see Figure 2.6). Each sensor features an infrared diode and a sensor unit that tracks the changes in infrared light at a distance of 10-80 cm. Users are recognized when one or two sensors are activated. Based on the physical distance of the

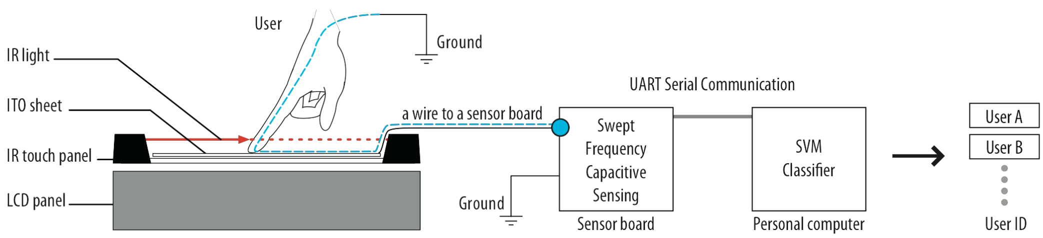


Figure 2.5: With an ITO-film, the user's capacity profile is measured and by the use of SFCS combined with a SVM mapped to a specific user. This figure was taken from Harrison *et al.* [2012b].

sensors, they assumed that only two sensors that are near to each other can be activated by one user.

Walther-Franks and Schwarten [2008] describe a similar approach and also use infrared proximity sensors to track and recognize users around an interactive tabletop. Similar to Tănase, the sensors are also places and the outer edge of the tabletop.

Evaluation: Both approaches recognize multiple users around a tablet and feature simulation input with multiple touch points. However, there is no direct mapping of a touch point to the user (I1 not fulfilled). Furthermore, the number of users is, in theory, unlimited, but is, however, bounded to the number of proximity sensors (I2). Nevertheless, users can freely position themselves around the tabletop and are not restricted to use additional device on their body (I4 and I7). Users lose their identification if they leave the sensor area (I6 not fulfilled). Positively mentioned can be the compactness of the system that can be integrated in mobile devices (I5). Robustness has not been analyzed by Tănase *et al.* and Walther-Franks *et al.* The systems itself do not need to be re-calibrated often; the recognition stays stable if the sensors are kept clean (I3).

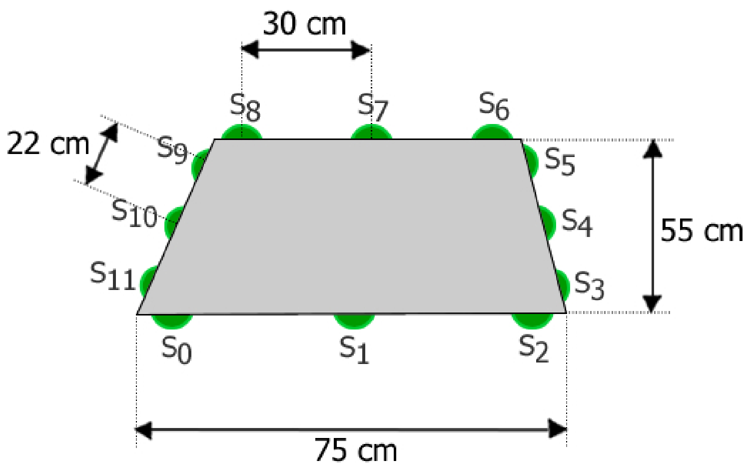


Figure 2.6: Twelve proximity sensors mounted onto the frame of a tabletop. Figure was taken Tanase *et al.* [2008].

Medusa

As a first proximity-based system developed [Annett *et al.*, 2011], Medusa associates the touch point to the user. Medusa has three different ring layouts (inner, outer, and outward ring) of proximity-based sensors (see Figure 2.7). Inner ring is used to track the user's hand, outer ring is for tracking the user's arm, and outward ring is used to recognize where the user is standing. Touch position provided by the screen combined with the sensors' data such as hand, arm, and user position allows a mapping of touch points to the specific user.

Evaluation: Medusa features a modern multi-touch technology and can therefore track multiple touches simultaneously (I1). Because of the high amount of sensors, there is no limitation regarding number of users (I2). Annett *et al.* mentioned that proximity sensors sometimes spuriously track approaching objects without having an object that is approaching the sensor, which leads to tracking errors (I3 partially fulfilled). This can be improved by the use of mirrors that, however, limit mobility and easy deployment (I5 partially fulfilled). Users are not augmented with additional devices that can be considered positively (I4). An everlasting user identification is not given similar to other proximity-based approaches (I6 not fulfilled). Users can freely position themselves around the tabletop (I7).

2.2.2.4 Additional Objects

In this section, techniques are presented for identification of the user with mobile devices, accessories, or gloves.

Mobile Phones

Already in 1998, Myers *et al.* [1998] presented PDAs that were used for personalized input on a common monitor. Hereby, the PDAs served as a personalized remote control (here cursor) for the additional screen.

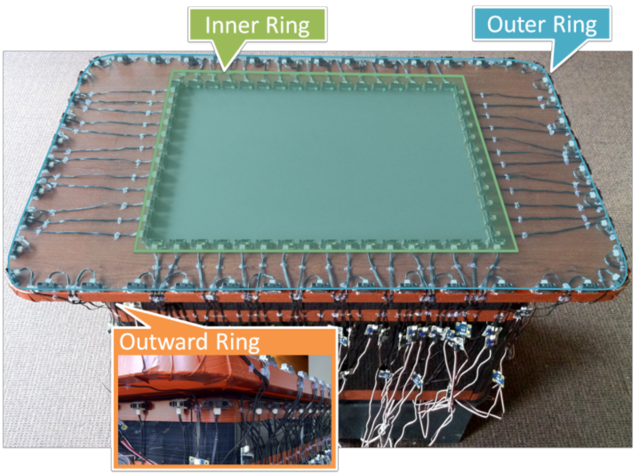


Figure 2.7: **Medusa has three different ring layouts of proximity sensors that are used to map the touch point to a user. Inner ring tracks the user's hand. Outer ring tracks the user's arm and outward ring tracks the user's position.**

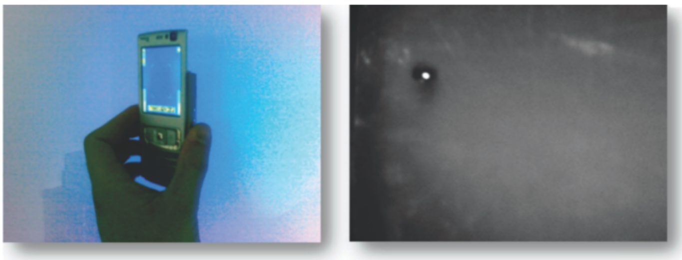


Figure 2.8: While placing a mobile phone on the touch display, it produces a pulse of light, which can be recognized by the camera image. This figure was taken from Schöning *et al.* [2008]

A similar approach was presented by Stewart *et al.* [1999] as well as Bier and Freeman [1991] that used multiple mice to provide personal input to the user (one mice per user). Schöning *et al.* [2008] presented a different approach to provide a user with personalized input on a touch screen. They used a mobile phone that was held onto the display and transmitted light pulse that was perceived by the touch display (see Figure 2.8). By the use of the mobile phone a user could specify a personal region for personalized work.

Evaluation: The system of Schöning, Rohs, and Kruger supports user identification; however, it is limited in simultaneously detecting the users. This technique, however, can either be used for interaction on the mobile phone or only in privately defined sub regions of the display (I1 partially fulfilled). Nevertheless, several users (I2) can be uniquely identified (I6) regardless of their position (I7). The use of an optical touch technology speaks in principle against a compact system design; modern techniques, however, allow a more compact optical recognition approach. Here, optical sensors are directly mounted on the display. Therefore, mobile use and easy deployment are quite conceivable (I5 partially fulfilled). The combination of identification number and light pulse can also guarantee a robust identification (I3). The use of a mobile phone, however, instruments the user (I4 not fulfilled).

Accessories

In addition to mobile devices, other accessory devices that are worn by the user were used to identify users. An example of one such accessory is the *IR ring* presented by Roth *et al.* [2010]. This small unit worn on the underside of the hand transmits via an IR diode, a unique light signal to the display. A similar approach of user identification with accessories worn by the user is *IdWristbands* [Meyer and Schmidt, 2010] featuring two IR diodes on a wristband worn by the user. In contrast to the IR ring, two IR diodes allows also to calculate hand orientation.

Evaluation: Both approaches, however, are only suitable for systems that use a camera for touch detection. However, this has a negative effect on the mobility of the system, since optical touch displays need considerably more space (I5 partially fulfilled). Furthermore, wearing of the IR ring or IdWristbands is necessary, and it can burden the user, because he will need to put it on before interacting with the system (I4 not fulfilled). In addition, a special hand pose and the long duration for a correct identification can have a negative effect. This is, for example, 2.4 seconds for authentication probability of 99.7%. The robustness of the method is thus dependent on the selected authentication probability (I3). Nevertheless, both systems allow a unique identification (I6 partially fulfilled) and free positioning of the user (I7) as well as provide simultaneous input (I1) for multiple users (I2).

Gloves

Although mobile devices and accessories feature user identification, they cannot distinguish different fingers of the user from each other. Therefore, Marquardt *et al.* [2010] have developed a system in which the users have to wear a special glove that has markers on each finger of the user. An optical tabletop approach can hereby distinguish the different 2 x 2 cm markers. A several-minute calibration process is needed so that the system learns and can identify different users.

Evaluation: The glove approach enables simultaneous and robust detection of the hands of several users (I1, I2, and I3). However, users are wearing gloves and are instrumented (I4 not fulfilled). Furthermore, the assignment of a glove to a user is based solely upon an identity of the glove. When gloves are exchanged, thus the identity of the user also changes (I6 partially fulfilled). By the use of an optical tabletop system, mobility is limited (I5 partially fulfilled). The system allows, however, a free positioning of the user (I7).

2.2.2.5 Pressure Sensors

Pressure-based techniques for user identification are based on the assumption that users can be distinguished and identified by looking at the pressure that they exert on a surface. The following techniques use pressure sensors that can measure the pressure on a surface and convert it into electrical signals.

Smartfloor introduced by Orr and Abowd [2000] consists of floor plates that are equipped with pressure sensors. If a person steps on one of the floor plates, the floor measures the obtained pressure over time and can generate a specific pressure profile for each user. This profile can then be used to distinguish people based on their way to approach or standing near the tabletop. This system can be used to identify the user and their position.

Evaluation: Orr and Abowd achieve a recognition accuracy of 93% for recognition of user identity. For recognizing and distinguishing users' feet, the recognition accuracy drops to 75%. This indicates a robust detection of a personalized input (I3). However, they do not present any assignment of a touch point to a user, they only determine user position (I1 not fulfilled). In principle, the allowed number of simultaneous users directly depends on the number of pressure-sensitive floor plates (I2). The user, while walking on the *Smartfloor* is not instrumented (I4) and can move freely. The mobility and deployment of the pressure-sensitive floor plates are limited (I5 not fulfilled). A unique user mapping is possible (I6) while allowing position invariance (I7).

2.2.2.6 Cameras

Various methods have been introduced that operate via image analysis using infrared, colored, or depth sensing cameras. In the following, we present some of the related previous works.

Differentiation of Hands

Methods supporting the detection of users' hands have their origin in hand, finger, or generally in the object shape recognition. Jain *et al.* [1999] used the thickness of the hand and length and width of the fingers and the ratio of the palm to the fingers as characteristics to distinguish users' hands. [2009] use size of fingertips and their orientation to distinguish different hands. In contrast, Oden *et al.* [2003] used contour and shape of the hand. In addition to the previous features, Sanchez-Reillo [2000] operated on hand geometric measurements, for example angle of the finger gaps, to identify users. Boreki and Zimmer [2005] use the curvature of the contour and the average finger length and width, for the extraction of characteristics. Other works cater to the use of fingerprints to improve recognition accuracy [Holz and Baudisch, 2010] or the creation of so-called fingerprint user interfaces [Sugiura and Koseki, 1998], which can provide personalized input on the fingertip.

However, with differentiation of the hands, personalized input for tabletop systems is not provided, because the assignment to the user is missing. However, those methods can be combined with existing approaches for detecting the user position (e.g., proximity sensors), so that a personalized input can be generated.

Dohse et al.

Dohse *et al.* [2008] presented a hybrid approach consisting of an optical tabletop system with an additional color camera for personalized input. In addition to touch points available from the tabletop system, the mounted color camera identified all hands in the field of view of the camera. Hands were identified by color matching and binarization of the colored image. All contours found in the segmented binary image are compared with the detected touch points of the optical tabletop system. A touch event is only triggered when the touch point is inside a contour.

Evaluation: The technique of Dohse *et al.* allows simultaneous detection (I1). However, only hand position and thus not user positions are recognized. Therefore, a unique user mapping is not possible (I6 not fulfilled). Nevertheless, multiple users can be detected simultaneously (I2) from each side (I7) without further instrumentation of the user (I4). Similar to other optical technologies that have a camera, a mobile use is costly. A single camera, however, can be easily deployed (I5 partially fulfilled). Another disadvantage of the method is that the output of the display must have a constant color and intensity if possible, otherwise the segmentation of skin color no longer works reliably (I3 not fulfilled). In addition, the user interaction is limited by the low frame rate of the system, providing only 16 frames per second.

HandsDown

Schmidt, Chong, and Gellersen presented *HandsDown* [Schmidt *et al.*, 2010], a system that also uses a hybrid approach that can identify users based on their individual hand contour. The system is an optical tabletop, which has been further enhanced with an optional infrared camera. The camera, however, is only used to pursue hands to maintain the predetermined identity. The contour of the hands seen by the camera are processed and hand features are identified and used to distinguish different users.

Evaluation: By analyzing the complete camera image, *HandDown* allows a simultaneous detection of multiple users (I1). For a group of 15 people, a recognition accuracy of 99.8% was achieved (I2 and I3). Hands can also be monitored over a longer period of time and thus maintain their classification (I6). Because of a rotation-invariant feature extraction, users can interact from any position (I7). To achieve user identification, an unknown user have to first place his hand flat on the tabletop so that the system can analyze and extract individual characteristics. This, however, is not natural (I4 partially fulfilled). The mobility of the system is due to an optical approach limited. The system is however easily to deploy (I5 partially fulfilled). Furthermore, the system can be influenced by large amounts of infrared light because of the use of an infrared camera that operate in the infrared light spectrum. Therefore, it is not suitable for outdoor use.

Carpus

In contrast to the techniques that allow differentiation of pure hands, in *Carpus*, introduced by Ramakers *et al.* [2012] the texture of the hand is used to extract bio metric features and assign a hand to a unique user. This approach allows not only a differentiation but also an authentication of the user by the use of available information in the user's hand.

Evaluation: The Carpus method is an improvement over *HandsDown*, because hands do not have to be placed flat on the table to be recognized. This reduces considerably the instrumentation of the user (I4). However, this method is sensitive to different lighting conditions and skin types. In optimal conditions, recognition accuracy is 97.3% for 20 registered users (I2 and I3). This value deteriorates to 87% when people are present who are not familiar with the system. Furthermore, *Carpus* allows the simultaneous detection of multiple users, while even allowing to differentiate between a left or a right hand (I1). The system needs an external camera that is easy to deploy. Mobile use of the system is, however, only limited possible (I5 partially fulfilled). Apart from the system disadvantages, the user association for personalized input is always maintained even if a user leaves the camera field of view (I6 and I7).

Bootstrapper

In contrast to the method explained above, Richter *et al.* [2012] present a system that is built on works from Orr and Abowd [2000] and Augsten *et al.* [2010] that recognizes and distinguishes users based on their shoe profile. They present *Bootstrapper*, a system that identifies users by their shoes. Hereby depth sensing cameras are mounted on all four sides on an tabletop to track and distinguish users' shoes (see Figure 2.9).

Evaluation: Bootstrapper can differentiate shoes of 18 different users with a recognition accuracy of 95.8% (I2). The precision of the association of users to touch points depends on the number of users. Consequently, the accuracy of the simultaneous detection of the personalized touch points varies from 99.7% for two users up to 92.3% for five (I1). Both accuracies speak for a robust system in terms of spontaneous interaction at a table , which, however, requires a flat placed shoe on the ground (I3 partially fulfilled). Furthermore, a little intrusive feature such as shoe is selected to differentiate users , which

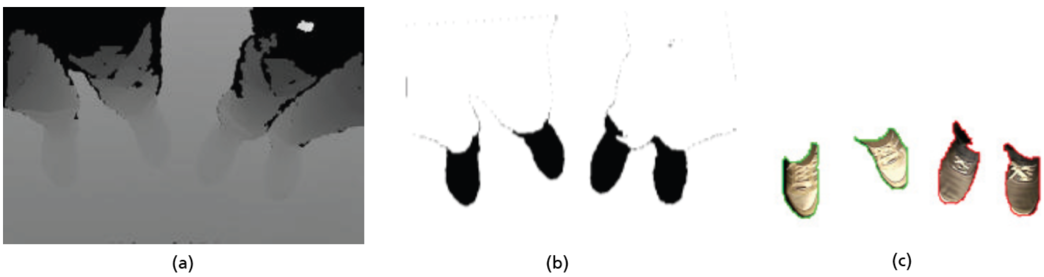


Figure 2.9: Bootstrapper can distinguish different users by recognizing their shoes with depth sensing camera. This figure was taken from Richter *et al.* [2012]

definitely cannot permanently differentiate users (I6 partially fulfilled). It allows for sufficient differentiability for the period of collaboration (I4, I7). The use of optical touch technology and four depth cameras disabled mobility and easy deployment (I5). In contrast to many other optical methods, however, the four cameras can be built into the table, thus allowing a more compact design.

Fiberio

Only recently, a new approach for personalized input was presented by Holz and Baudisch [2013] called *Fiberio*, which is a rear-projected tabletop system that identifies the user by identifying her fingerprints. By the use of a fiber-optic plate, (1) back-projected light can be seen by the user from all locations around the table and (2) IR light can be reflected in a specific way so that fingerprint ridges and valleys become visible for a high-resolution infrared camera mounted below the table (see Figure 2.10).

Figure 2.10: *Fiberio* allows user identification and authentication by a precise recognition of their fingertips. This figure was taken from Holz and Baudisch [2013]

Evaluation: Fiberio allows simultaneous input (I1) for multiple users (I2). The system works with user fingerprints that can be considered reliable for user association (I6). The system does not need to be re-calibrated (I3). Users can freely interact with the system without any instrumentation (I4) while also freely position themselves (I7) around the table. Because Fiberio uses a rear-projection approach and needs rear high-resolution infrared camera, mobility is limited. The deployment, however, is easily possible (I5 partially fulfilled).

2.2.2.7 Comparison of Techniques for Personalized Input

Previously presented methods are compared. The requirements defined in subsubsection 2.2.2.1. Therefore, the presented systems are measured in relation to these in Table 2.1.

2.2.3 Personalized Output

In this section, we will present previous research that focused on providing each user with an own output (in the following, we call it *personalized output*).

This can be supported when each user has the personalized output on an additional mobile device [Seifert *et al.*, 2012; Sugimoto *et al.*, 2004; Wigdor *et al.*, 2009] (see subsection 2.2.1 for a detailed overview of related work in this area). However, all these approaches require the collaborators to *switch their attention* between the surface and the secondary device.

With this in mind, we are reporting on research projects that tried to support so-called *multi-output* displays. Multiple outputs are provided with a single display. Each user can simultaneously see different content (personalized output) on the very same screen (see Figure 2.11). These displays use different technologies that can be divided into systems that use or do not use specific glasses. Figure 2.12 shows an overview of existing techniques. In the following, in subsubsection 2.2.3.1, we first explain the requirements defined for systems with personalized output. Then in subsubsection 2.2.3.2, we present stereoscopic systems that require special glasses. In subsubsection 2.2.3.3, autostereoscopic systems are explained that work without glasses. Finally, the explained techniques are compared in subsubsection 2.2.3.4.

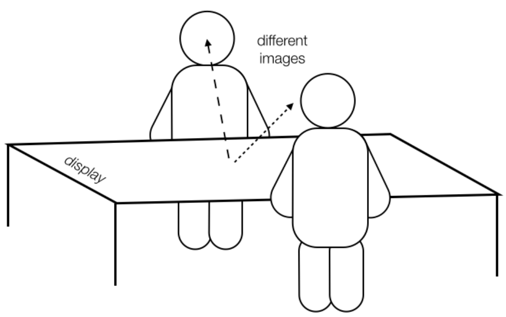


Figure 2.11: Multiple outputs are provided with a single display. Hereby each users sees different content on the same display simultaneously.

2.2.3.1 Requirements for Personalized Output

In order to compare previous related work on personalized output, we have identified requirements for personalized output that will be presented in the following.

	Requirements for Personalized Input	Simultaneous input I1	Multiple users I2	Robustness I3	No User Instrumentation I4	Easy Deployment I5	Sustain. User Association I6	Free User Positioning I7	Remarks
Capacity	DiamondTouch [Dietz and Leigh, 2001]	●	●	●	●	○	◐	○	1.
	Capacitive Fingerprinting [Harrison *et al.*, 2012b]	○	●	●	●	●	●	●	2.
Proximity	Tănase *et al.* [Tanase *et al.*, 2008]	○	●	●	●	●	○	●	
	Walther-Franks [Walther-Franks and Schwarten, 2008]	○	●	●	●	●	○	●	
	Medusa [Annett *et al.*, 2011]	●	●	◐	●	◐	○	●	
Additional Objects	Mobile phones [Schöning *et al.*, 2008]	◐	●	●	○	◐	●	●	
	IR Ring [Roth *et al.*, 2010]	●	●	●	○	◐	◐	●	
	IdWristbands [Meyer and Schmidt, 2010]	●	●	●	○	◐	◐	●	
	Groves [Marquardt *et al.*, 2010]	●	●	●	○	◐	◐	●	
Pressure	Smart Floor [Orr and Abowd, 2000]	○	●	●	●	○	●	●	
Cameras	Dohse [Dohse *et al.*, 2008]	●	●	○	●	◐	○	●	3.
	HandsDown [Schmidt *et al.*, 2010]	●	●	●	◐	◐	●	●	
	Carpus [Ramakers *et al.*, 2012]	●	●	●	●	◐	●	●	
	Boostrapper [Richter *et al.*, 2012]	●	●	◐	●	○	◐	●	
	Fiberio [Holz and Baudisch, 2013]	●	●	●	●	◐	●	●	

Table 2.1: Personalized input techniques. ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

Remarks: 1. Users have to sit on their chairs during interaction. This is considered as a disadvantage. 2. This approach interferes with capacitive touch technology. 3. User's interaction is limited by the low frame rate of the system providing only 16 frames per second.

O1: Scaleable Simultaneous Output Obviously, the availability of more than one output is crucial for a personalized output. Multiple output needs to be provided simultaneously so that users can work in parallel. This requirement assesses how scaleable is the provided solution.

O2: Overlap of Private and Group Territories [Scott *et al.*, 2004] The location and size of the private and group territory changes, depending on the used techniques. This may lead to an overlapping of private and group territories.

O3: Robustness The simultaneous output to multiple people should be robust. The system should maintain its functionality also under bad light conditions.

O4: No User Instrumentation Similar to personalized input, personalized output can also instrument the user. This should be avoided to allow spontaneous natural interaction with the system.

O5: Mobility It is useful to develop techniques that have a certain mobility. This includes, for example, a compact design that can be integrated into mobile devices.

O6: Free User Positioning The system should support a free positioning of the user around the tabletop, thereby providing personalized output independently of the viewing angle of the user.

2.2.3.2 Stereoscopic Systems

Nowadays, the stereoscopic approach is often used in 3-D TVs to allow users to view 3-D movies. Stereoscopic displays work with the following principle: The display shows in short intervals a minimum of 120 Hz (faster than what our human eye can see, 60 Hz) different images. Each lens of the glass switches every 60 Hz between two states, transparent and opaque, respectively. For 3-D TV images are alternately shown to the left or the right eye of the user. With high speed, our brain combines the images to a 3-D

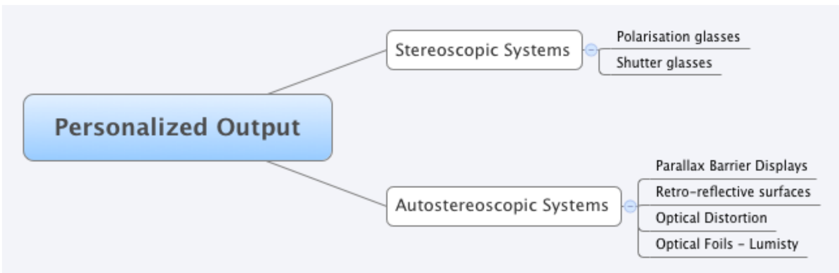


Figure 2.12: Different techniques for personalized output that will be presented in the following.

image. In the following related works the stereoscopic approach is misused to allow each user to see a different image on the whole screen. Both lenses of the glasses, as opposed to only one as that for 3-D TVs, are switched between the transparent and the opaque state.

The use of stereoscopic systems always requires special glasses. These are based either on the shutter or the polarization principle. First, systems are presented that work with the shutter glasses. Then, the polarization technique is explained in more detail. Through the use of glasses to allow personalized output, the user is always instrumented to a certain degree (O4 not fulfilled).

Shutter Glasses

Shutter glasses consist of two liquid crystal lenses that have two states, either transparent or opaque. Both states can be switched by adding or removing current to the liquid. This allows to block or transmit the light to the user emitted from the display (see Figure 2.13).

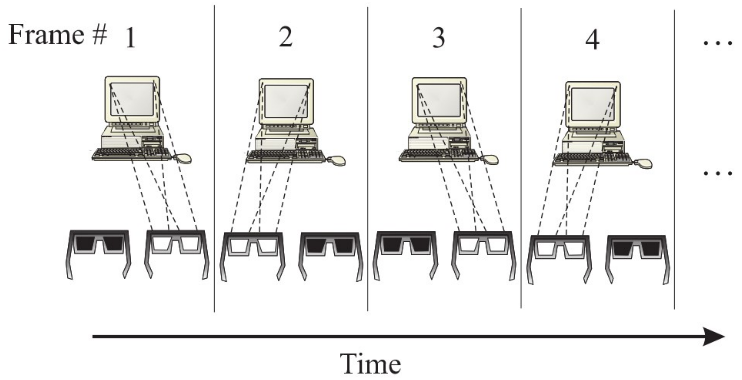


Figure 2.13: In the first time step, the glasses of the left and the right glasses are transparent and the right glasses are opaque, so that the image shown on the display is visible for the second user. Then both glasses switch their state allowing only the left user to see the screen. This state switch iterates over time. Each time step is 60 Hz. This figure was taken from Shoemaker and Inkpen [2001].

Two-User Responsive Workbench

In 1997, Agrawala *et al.* introduced already *Two-User Responsive Workbench* [Agrawala *et al.*, 1997]. The system allows two users to simultaneously view through glasses two stereoscopic images that show a virtual reality. Users see the same virtual world, however, from a different viewpoint.

Evaluation: They used the shutter technique to personalize the output for two users

(O1). They were able to show 3-D content to each user by featuring four images that could be displayed one after the other. Simultaneously, only one glass per user is opened, so that a single image could be sent to each user's eye. By tracking the head position, an own 3-D perspective of a scene was provided to each user. This allowed coverage of personalized and common work area (O2) and a free position of the users around the scene (O6) . However, the Two-User Responsive Workbench has only a refresh rate of 120Hz, four separate images are used simultaneously it leads to a refresh rate of 30 Hz per image, which causes a flickering that was considered physiologically stressful (O3 not fulfilled). Moreover, the use of a projector system to visualize the content and additional techniques for detecting the head position restrict mobility of the system. (O5 not fulfilled).

thirdEye, Shoemaker, and C1x6

Another project, *thirdEye* [Mistry, 2009], also uses shutter glasses in companion with an LCD display that allows a user personalized output with an appropriate refresh rate of 60 Hz. Shoemaker and Inkpen [2001] also use shutter technology to provide personalized output for an external monitor. Likewise, C1x6 developed by Kulik *et al.* [2011], who based their technique also on the shutter technique.

Evaluation: The authors use projectors with a very high frame rate of 360 Hz to allow simultaneous use up to six users (O1), regardless of their position (O6). In contrast to the system of Agrawala *et al.*, Kulik's frame rate is 60 Hz per person (O3). This value can be adaptively improved when the system would dynamically adjust the frame rate depending on the number of users. Furthermore, by using differently polarized images, three-dimensional content could be provided. Becaues of the focus on large screens, mobility is limited (O5 not fulfilled). Overlap of group and private workspace is possible on the whole screen (O2).

Temporal Psycho Visual Modulation

A conceptual paradigm called *Temporal Psycho Visual Modulation* (TPVM) presented by Wu and Zhai [2013] extends the shutter technology with a physiological component by explaining the psychophysics of human vision. Human visual systems function with approximately 60 Hz for most viewers. Humans cannot resolve flicker images that go beyond this flicker frequency of 60 Hz. Based on this knowledge, he presents a conceptual calculation saying that if display refresh rates increase, the number of users that can see different content on a screen also increase.

Evaluation: With the TPVM technique, it is conceptually possible to implement a common view that is visible to all users without glasses and, if required, having personalized information (by wearing appropriate glasses) (O2). Since TPVM is not bound to a special tabletop technology but only increased refresh rate, the concept would also work for mobile devices (O5). Similar to other shutter techniques, the user is indeed instrumented (O4) and is free to choose her position (O6). By the use of a nonnegative matrix decomposition of the frame rate an optimized and more robust representation of content is possible (O3).

Polarized glasses

So-called polarized glasses, which use the principle of polarization of light, allow also personalized output. Light consists of an electromagnetic wave that can be polarized in a linear or nonlinear way. In order to personalize user's output, the light source (e.g., LCD display) has to emit in different polarization. In addition, users' glasses need appropriate *polarizers* (also known as polarization filters), which pass only the polarized light of the particular user view.

Sakurai *et al.* [2008, 2009] used the polarization filters to provide a user with personalized output. In their system, own elements appear brighter than the foreign ones. The authors use two projectors. One projector's light is linearly polarized with a polarizer (see Figure 2.14). The polarizer can be rotated 360°. Each user wears polarized glasses. The information visibility is determined by the user's viewing angle and the polarized light source. With the knowledge of user positions, projected content can be privately shown to a user, when the rotated polarizer is aligned with the user's position. All other users will not or only slightly see this information due to their different points of view.

Evaluation: Thus, with the use of this system, it is more likely to see other users' information (O2). A further problem is the use of linearly polarized light. If users are near to each other or turn their head so that they have the same viewing angle, they would completely see other users' view (O3). There solution also limits the user to choose a free positioning around the tabletop (O6 partially fulfilled) . With modern technology, it is possible to avoid the previously described effect of the use of circularly polarized light. Likewise, a more compact design with projectors are possible in modern systems (O5 not fulfilled). Furthermore, the number of simultaneous outputs dependent on refresh rate of the projector, synchronization speed of the projector, and orientation of the polarizer. Sakurai *et al.* does not, however, report these details about their system (O1). Here, polarized glasses are passive as opposed to active shutter glasses that need battery. Polarized glasses are thus on average cheaper and lighter than shutter glasses. However, the user is also in this form of personalized output instrumented (O4).

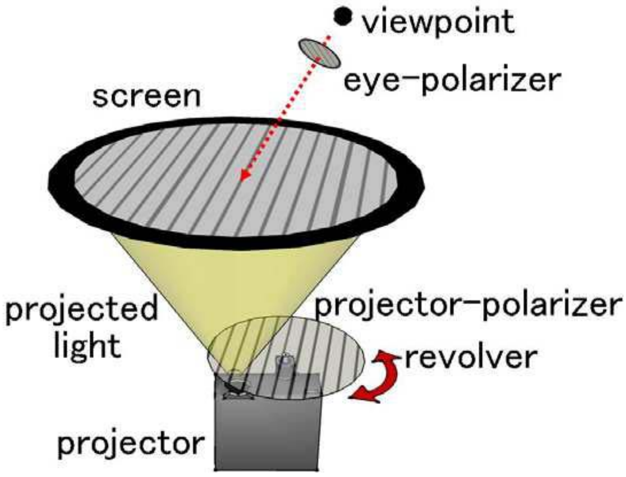


Figure 2.14: The light from the projector passes a rotatable polarizer so that it can only be perceived from a particular viewing position available via polarizing glasses. This process is repeated for all users and their position in a very short time and thus allows a personalized output for each user. This figure was taken from Sakurai *et al.* [2008]

2.2.3.3 Autostereoscopic Systems

In contrast to the previously discussed stereoscopic display techniques, *autostereoscopic displays* allow personalized output without glasses. This advantage is based on the fact that different users also are mostly at different positions in space with a different view angle onto the display, which can be used to provide personalized output. Different approaches are presented in the following in more detail.

Parallax barrier displays

Parallax displays use the fact that different users often have different positions around the display leading also to a different viewing angle. A display can, in front of it, be augmented with vertical grating (*parallax barrier)* or *lenticular lens array*. With a different angle of the user to the display, different information is visible to the user (see Figure 2.15).

Lenticular displays use many small lenses that are placed in front of the pixels of the display. This refract the light emitted by the pixel via the lens in different directions. Similarly, parallax displays have so-called parallax that block the light from the pixel of the display in a certain direction. For both techniques, the display needs to be viewed from a certain point of view to display different content. In their initial realization, this forced the user to stay at a fixed position all the time. There are variations, however, that can compensate this disadvantage by using dynamic parallax.

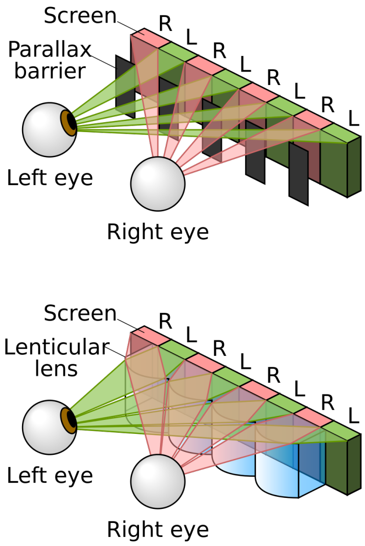


Figure 2.15: Comparison of (top) lens array and (bottom), parallax barrier displays. Person A, indicated as a left eye, sees due to the refraction of the lens (bottom) or blocking by parallax barriers (top) only the greens, pixels whereas person B, indicated as right eye, sees only red ones. This figure is based on this source `http://commons.wikimedia.org/wiki/File:Parallax_barrier_vs_lenticular_screen.svg`, last check: 03.12.2013)

Perlin

In 2000, Perlin *et al.* [2000] have developed an autostereoscopic display using an parallax barrier display with dynamic barriers (also called *dynamic parallax*). Dynamic parallax allows to switch the barriers between two states: transparent or opaque. This allows to display personalized content also when users move their viewing position, which was not possible with fixed barriers. Similarly, Peterka *et al.* [2007] developed a so-called dynallax display by using liquid crystal display pixels, providing four different views. The combination of user tracking with an InterSense motion tracker [InterSense, 2006] and dynamic parallax maximize efficient use of the number of existing pixels. This approach, however, decreases the brightness of the display, because light is partially blocked while passing trough parallax barriers. Parallax display have a restriction in viewing direction. They can only be viewed from horizontal or vertical point of view. For a tabletop , which needs to be accessed from all sides, parallax technique is not suitable.

IllusionHole

In this domain, Kitamura *et al.* have presented *IllusionHole* [Kitamura *et al.*, 2001, 2005] that is a parallax barrier display that works when users are seated around the display (see

Figure 2.16) . Users view the display through a circular mask in front of the actual display. Because of the different horizontal or vertical directions, angle users can see a different section of the underlying displays. At the same time, the area directly beneath the circular hole can be seen as a common interaction surface. Different users were tracked by a Polhemus Fastrack (magnetic tracker) or an Intersense IS-600 Mark 2 motion tracker. In addition Kitamura *et al.* used shutter glasses and users pose to provide three-dimensional content.

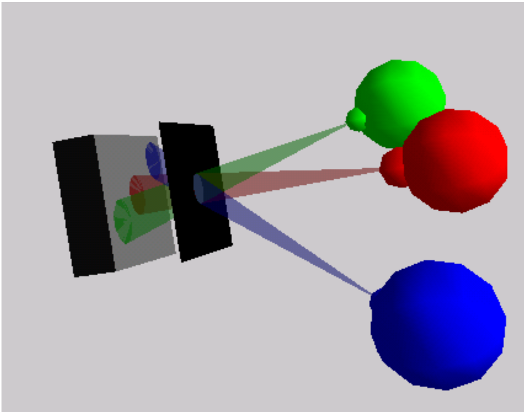


Figure 2.16: Because of the circular mask, users with a different view angle (red, green, blue) can see a different part of the underlying screen and therefore perceive personalized content. However, the range of different position is limited. This figure was taken from Kitamura *et al.* [2001].

Ye

Ye and Fuchs [2010] modified the parallax barriers in a similar circular way as Illusion-Hole. The resulting tabletop display allows a personalized viewing from any direction around the display by tracking user positions with a marker-based approach. Display pixel are then dynamically changed based on the user position. Matusik *et al.* [2008] introduced a similar tabletop system limiting the user with a fixed position in space. Smith and Piekarski [2008] introduced in also a tabletop system allowing up to four views by the use of liquid crystal display. However, the display is not really usable because, as reported by Smith, the text becomes unreadable.

In contrast to parallax displays, lenticular displays are brighter, since the lenses pass trough more light. However, lenticular techniques only provide personalized output from a horizontal view point direction and are therefore not suitable for tabletop systems. Lincoln *et al.* [2009] used a lenticular approach to implement a video conferencing system that allowed two participants to see correct perspective of the remote collaborators.

Evaluation of lenticular and parallax barrier displays: Lenticular and parallax barrier displays allow a potentially large number of simultaneous output (O1). However, the resolution per user is correspondingly reduced by each additional user. Furthermore, the systems allow overlapping of private and common group work area (O2) and is suitable for mobile use, because the techniques can be implemented on pixel level (O5). Many systems allow a free positioning of the user (O6), this needs a recognition of the location of the user, which then restricts the mobility of the system. Lenticular and parallax displays do not require to instrument the user (O4). If several users are close to each other, it most probably can happen that the view of the other user is visible, because these techniques are highly dependent on the user's point of view (O3 partially fulfilled).

Retro-Reflective Surfaces

While using *specular reflection*, similar to a normal mirror, the angle of light incidence equals the angle of reflection (incoming angle of light = outcoming angle of light). A *retro-reflective surface* consist of a so-called retro-reflector. The specialty of this material is that incoming light is reflected back in the same angle of entry (see Figure 2.17). Such materials are often used in traffic control to increase visibility of traffic signs. For example, mirrors are using

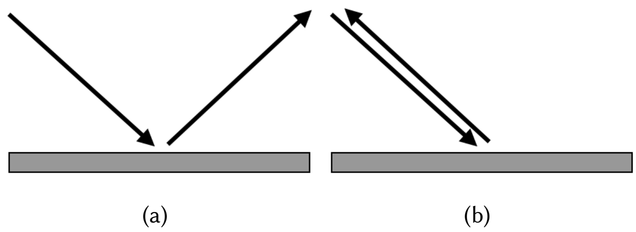


Figure 2.17: Comparison between specular reflection (a) and retro-reflection (b). In contrast to specular reflection, retro-reflection reflects the incoming light source back with the same angle of entry.

Hua and Brown [2003] introduces a system called *SCAPE* that uses retro-reflective surfaces. A projector is mounted on the user's head. The projector is used to visualize private information. If a user lifts a retro-reflective surface in front of his projection (here his point of view), private information is visualized on it, because the material refracts the projection only in the user's direction.

Evaluation: Obviously, *SCAPE* needs an enormous instrumentation of the user by featuring a head-mounted projector that needs cable to the power supply (O4 not fulfilled). However, it allows personalized output for each user (O1) with free head movement and no limitation on user position (O6). Private content, however, cannot be overlapped with group content because the system only has support to visualize private content. Group content is also privately reflected to only one user. (O2 not fulfilled). Despite the simplistic approach, the private content can be easily viewed by moving a hand or another non-retro-reflective object into other users' projection (O3 not fulfilled). The system can be used in mobile setting because it only needs a projector and retro-reflective surface (O5).

Optical Distortion

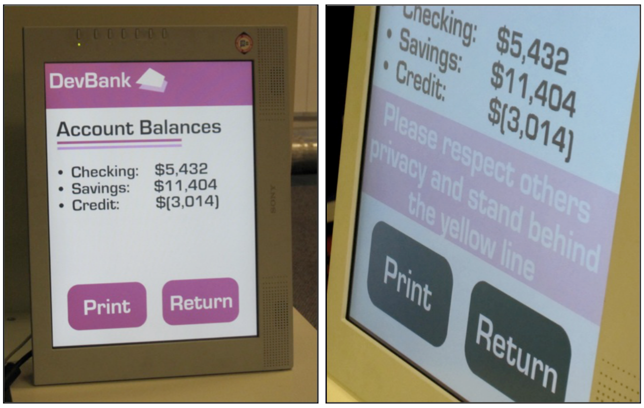
The term *optical distortion* in the following should be understood as direction-dependent optical effects with liquid crystal displays. Harrison and Hudson [2011] used the phenomenon on cheap liquid crystal displays that colors vary dependent on the horizontal viewing angle. Because of this property, special colors can be used that are visible or invisible depending on the specific viewing angle (see Figure 2.18).

Harrison

Evaluation: Optical distortion effects occur, according to Harrison and Hudson, only in the horizontal orientation. Thus, the method is limited to two simultaneous outputs (O1) and is strongly bound to users' viewing position (O6 not fulfilled). Furthermore, the system allows mobile use (O5) and does not instrument the user (O4). The range of possible colors is, however, limited. Therefore, this method is only suitable for a very careful design of specially colored content (O3 not fulfilled). A presentation of personalized content and shared content in the same place is possible, but it needs a strong restriction in the displayed content because of high restriction in colors that need to be used (O2 not fulfilled).

Kim

Evaluation: Kim *et al.* [2012] develop a similar approach as Harrison that provide personalized content depending on the viewing angle of liquid crystal displays. He investigated the brightness curves and varied the viewing angle to understand the colors that need to be used to visualize different content depending on the viewing angle. Through the optimization of the contrast, Kim *et al.* allowed two simultaneous displays (O1) to have overlaps to some content (O2 partially fulfilled). Similar to Harrison the user is severely restricted in the viewing angle (O3 and O6, both not fulfilled) and is not instrumented (O4). Furthermore, a mobile use with existing technology is possible (O5). The implementation as a tabletop system is conceivable; however, leads to problems, since this approach only works and changes content depending on the change in horizontal viewing direction.



(a) (b)

Figure 2.18: A liquid crystal display can have additional personalized information for a second user (b) which, due to the optical distortion of the user point of view, is only visible to one of them (a). This figure was taken from Harrison and Hudson [2011].

Optical Foils-Lumisty

In contrast to optical distortion, which is a side effect of using liquid crystal displays, optical foils exist that can get opaque or transparent to the users eye depending on the viewing angle.

Lumisight

The *Lumisight Table* by Matsushita *et al.* [2004] is a tabletop system that has a special optical foil called *Lumisty*. This film has the ability to get transparent or opaque depending on the viewing angle (see Figure 2.19) . This figure shows that a light source can pass through the foil of the film from an angle $-25°$ to $25°$. The same light source is diffused from a viewing angle of $-25°$ to $-55°$ or $25°$ to $55°$. This effect is used by the authors by placing projectors underneath the table and setting up the projectors angles in a way that different users with different viewing angles can only see their projected image. This can be done for four views (O1), by laying two slides of orthogonal Lumisty foils over each other. Input is done with tangible objects that are tracked on the surface. Tangibles block the projected content that is why the concept of Kakehi *et al.* [2006] extended this setting with transparent tangible objects.

Evaluation: The authors use a Fresnel lens in order to keep size of the system as compact as possible. Through the use of one projector per user, the mobility of the system is limited (O5 not fulfilled). The user is not instrumented (O4), and user position/viewing angle is

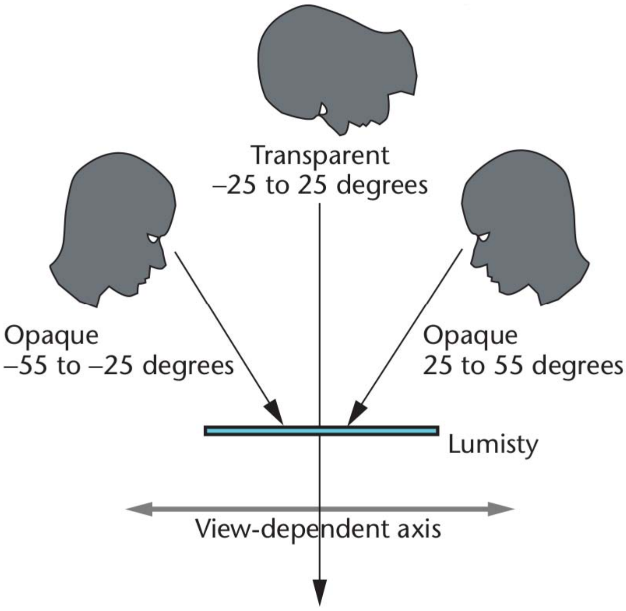


Figure 2.19: Depending on the viewing angle, the lumisty foil can be transparent or opaque. This figure was taken from Kakehi *et al.* [2005].

strongly tied to a sitting position and needs to be predetermined by the arrangement of the projector's position. The user position may differ only by 10° from the predetermined position; otherwise, the personalized output of users become visible (O6 not fulfilled). Nevertheless, *Lumisight Table* allows an overlapping private and group workspace on the entire surface (O2) that is not influenced by the light conditions (O3 partially fulfilled).

Kakehi, Honda

Another similar work called *UlteriorScape* developed by IEEE [2008] is a refinement of the Lumisight Table. This tabletop system additionally offers personalized views that are tracked above the tabletop surface. Honda and Sakamoto [2010] also developed a tabletop system for four users with personalized output, which uses Lumisty foil in combination with liquid crystal displays. Evaluation results for this systems are the same as for Lumisight.

2.2.3.4 Comparison of Presented Techniques for Personalized Output

The previously presented techniques for personalized output are compared with respect to the defined requirements in subsubsection 2.2.3.1. The comparison is presented in Table 2.2.

	Requirements for Personalized Output	Scaleable Simul. Output	Overlapping	Robustness	User Instrumenta.	Mobility	Free Positioning	Remarks
		O1	O2	O3	O4	O5	O6	
Stereoscopic Systems	Two-User Responsive Workbench [Agrawala *et al.*, 1997]	2	●	○	○	○	●	1.
	thirdEye [Mistry, 2009]	2	●	●	○	●	●	1.
	Shoemaker *et al.* [Shoemaker and Inkpen, 2001]	2	●	●	○	●	●	1.
	C1x6 [Kulik *et al.*, 2011]	6	●	●	○	○	●	1.
	TPVM [Wu and Zhai, 2013]	4	●	●	○	●	●	1.
	Sakurai [Sakurai *et al.*, 2008]	2	●	●	○	○	◐	1.
Autostereoscopic Systems	Perlin [Perlin *et al.*, 2000]	2	●	◐	●	●	●	
	Dynallax Display [Peterka *et al.*, 2007]	4	●	◐	●	●	●	
	IllusionHole [Kitamura *et al.*, 2001]	4	◐	◐	●	●	○	
	Ye [Ye and Fuchs, 2010]	4	●	◐	●	●	●	2.
	Matusik [Matusik *et al.*, 2008]	2	●	◐	●	●	○	
	Smith [Smith and Piekarski, 2008]	4	●	◐	●	●	○	
	Lincoln [Lincoln *et al.*, 2009]	2	●	◐	●	◐	○	
	SCAPE [Hua and Brown, 2003]	n	○	○	○	●	●	
	Harrison [Harrison and Hudson, 2011]	2	○	○	●	●	○	
	Kim [Kim *et al.*, 2012]	2	◐	○	●	●	○	
	Lumisight Table [Matsushita *et al.*, 2004]	4	●	◐	●	○	○	

Table 2.2: Comparison of presented techniques for personalized output. ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

Remarks: 1. The use of stereoscopic systems always require special glasses. 2. The resolution per user is reduced by each additional user.

2.2.4 Private Information Space

Personalized visibility and accessibility for each user is provided by the following tabletop systems, allowing each user to have his own private information space with personalized input and output. Consequently, these systems need to match both requirements for personalized input (see subsubsection 2.2.2.1) and personalized output (see subsubsection 2.2.3.1).

2.2.4.1 PiVOT

PiVOT [Karnik *et al.*, 2012] is a tabletop system introduced by Karnik *et al.* that divides the table into two zones of vision:

(1) All users are able to interact on the common shared layer. A shared view is visible when users' are looking from the side at the display.

(2) Each user has a personalized layer that does not interfere with other users personal layer. This layer is shown to the user if she leans forward to view the display from above (perpendicular to the display surface). Only the owner of the respective personalized area is allowed to interact with it.

In order to implement the two different layers, Karnik *et al.* (similar to Matsushita *et al.* for Lumisight Table) used a Lumisty foil that is diffuse for a light source coming from an angle of $25°$ or $55°$ and transparent otherwise, so that a projected perpendicular light can pass unhindered. A projector from the side is used for a common shared layer for all users, which is visible from the side (see Figure 2.20) .

In addition, to the Lumisty foil, a so-called *LC-sandwich* is used through which the personal view is possible. The LC-sandwich is a data layer that shows the personalized information by using a mask (*Mask LC*) that has the same principle as in parallax displays. It is important to know that the data layer can pass unpolarized light unhindered. Thus, the projected content can unimpededly reach through the Lumisty foil. Polarized light, however, is dependent on the polarization direction that is controlled (blocked or forwarded) by the mask.

In order to allow a personalized output, all users have to wear markers on their heads that are used to detect users' position and head orientation. The mask is parallax from the approximated eye/head position so that every user sees his own personalized layer. It is also possible to display three-dimensional content on the personalized layer. PiVOT allows a personalized input via touch; it is unclear, however, how this functionality was

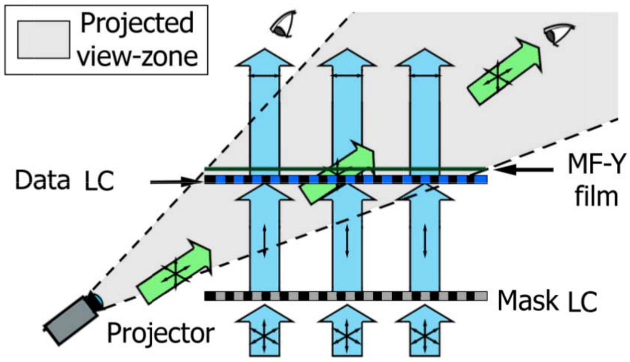


Figure 2.20: A projector is used from the side at a viewing angle of 25° to 55° to project a shared view on the Lumisty in their diffuse state (MF-Y film) because the light can freely pass trough the data layer (Data LC) as unpolarized light. The combination of the data layer with the controllable mask layer (Mask LC) allows a personalized output to the user while leaning forward. This figure was taken from Karnik *et al.* [2012].

implemented. Presumably, it is based on mapping detected touch points to the nearest user.

The the following, we discuss PiVOT under the defined requirements for in- and output in subsubsection 2.2.2.1 or subsubsection 2.2.3.1.

PiVOT can only be used from three sides of the screen, as the projector is mounted on the rear side of the table. Furthermore, the systems require a camera to calculate users' head position that limits mobile use. This approach, however, is easy to deploy. (I5 and O5 partially fulfilled) . The user does not need to wear glasses as for the stereoscopic techniques. It is still a form of instrumentation and change interaction with the system when users need to have markers on their head. Markers are needed for both personalized in- and output so that users are instrumented in relation to both requirements (I4 and O4 not fulfilled) . Through the use of markers, a free positioning on three sides of the table is possible (O6 and I7), allowing a unique assignment of a user, even if it leaves the sensor area (I6) . In addition, the number of users is directly dependent on differentiation of the markers (I2). An overlap of common and personalized workspace is possible; to change these, however, requires a change in the head position of the user (O2 partially fulfilled). The system offers several personal views (O1). An overlap of several personal layers when users look at the surface from the same view angle can cause a user seeing the other user's personal view. This can lead to problems with robustness (O3 partially fulfilled). Karnik *et al.* mention that personalized input is possible, the implementation, however, is unspecified. Therefore, statements about robustness (I3 partially fulfilled) and the ability to control multiple personalized layers via touch points simultaneously can not be made (I1 partially fulfilled).

2.3 Supporting Fluid Information Spaces

In the following, we describe our concept that provides support for fluid collaboration. In addition to the previously introduced technical requirements for personalized input and output, we first introduce requirements that should be met while supporting fluid collaboration on a multi-view tabletop. Second, we describe a concept that allows for flexible transitioning between coupling styles on a tabletop. Next, the technical realization of the concepts are presented. Last, we present example applications for illustration and evaluation of interaction and visualization techniques.

2.3.1 Requirements

In addition to the technical requirements such as personalized in- and output, requirements (R1-R3) focusing on collaboration on such multi-view tabletop are needed and are presented in the following:

R1: Support of same screen group and private views. Collaborators need to easily and seamlessly transition between both ends of the coupling spectrum (tight coupling vs. loose coupling). Group views are suitable for synchronizing the working state while having a common ground; the private views provide high-resolution personal workspaces, to conduct independent work unobtrusively, as recommended by Tang *et al.* [2006].

R2: Support of mutual awareness and coordination in private views. While group views provide mutual awareness by providing a common ground, working in independent private views on the same surface needs interaction techniques that support mutual awareness and coordination.

R3: Reducing interference in the group view. While collaborators can effectively work in parallel on content that is juxtaposed on the shared group view, overlapping content is problematic. The resulting occlusion is disruptive and can prohibit other collaborators from accessing occluded elements.

These requirements were key rationals for our design and evaluation of our concept and interaction techniques. These requirements are supported and verified by the interaction and visualization techniques (presented in section 2.4) in combination with the multi-view tabletop that supports personalized input and output.

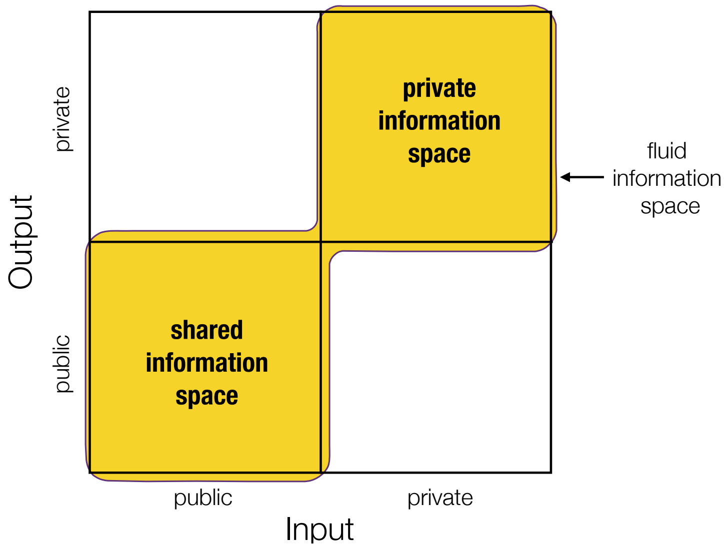


Figure 2.21: Augmented design space with the contribution of this chapter: Fluid information space, called Permulin, allows fluent transition between private and shared information space on a tabletop.

2.3.2 Concept

We present *fluid information space*, we further call Permulin, which allows fluent transition between private and shared information space on the very same tabletop (see Figure 2.21).

Permulin is an integrated set of novel interaction and visualization techniques for fluid collaboration on multi-view tabletops. Permulin allows to fluently switch between a full-screen group or private view, allowing distinct private views or a group view that is overlaid with private contents for each user. Our techniques provide support for easy, seamless, and gradual transitions on the entire spectrum between tightly coupled and loosely coupled collaboration (see Figure 2.22). This fluent switch is highly important while working in fluid collaboration, when users frequently switch between individual and group work.

In detail, Permulin provides:

(1) a *group view* for common ground during phases of tight collaboration,

(2) *private view* for each collaborator to scaffold loosely coupled collaboration, and

(3) interaction and visualization techniques to switch between private and group view individually and share content in-between these views for coordination and mutual awareness.

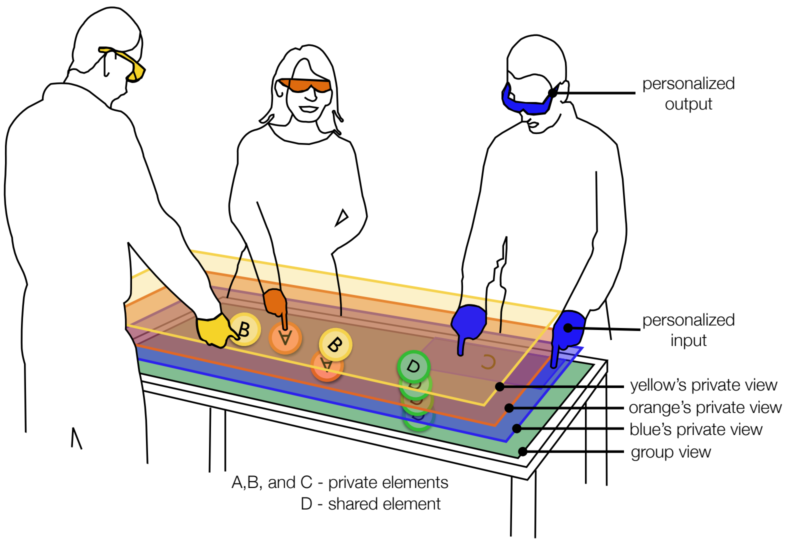


Figure 2.22: Conceptual visualization of Permulin. Each user has her personalized in- and output visualized by her own colored glasses and hands. Each user can switch between full-screen private or group view.

As pictured in Figure 2.22, Permulin conceptually provides a personalized in- and output to each user. Information that is visualized privately on the private view can only be accessed and manipulated from the corresponding user. Moreover each user can switch between full-screen private view and a full-screen group view as well as share content in-between these views with other collaborators, respectively.

As multi-view tabletop (see Definition 8) hardware seems highly promising to support fluent transitions between tight and loose coupling styles [Tang *et al.*, 2006] in fluid collaboration. Our Permulin system concept was built upon two main features: (1) *personalized output*, to ensure users can have full-screen private views that can overlap with common group view, and (2) *personalized input*, to have a one-to-one mapping between user interaction and his view on the multi-view tabletop. Personalized output can be implemented for multiple users using an stereoscopic approach combined with the use of active shutter glasses. The display and the glasses are synchronized to allow each user to see only his content, respectively. Personalized input can be provided by visually tracking and identifying users' hands by addition cameras.

As mentioned, Permulin contributes interaction techniques and visualization techniques for fluent mixed-focus collaboration. In the following, we elaborate in detail how we have technically realized our system.

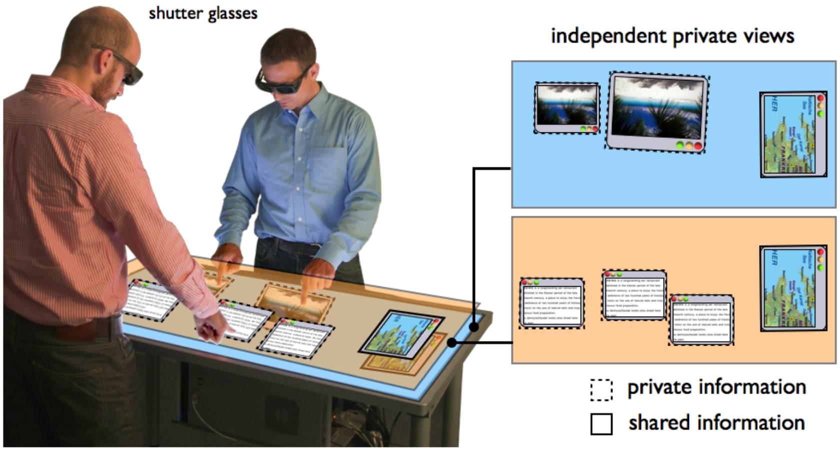


Figure 2.23: Current realization of the Permulin concept.

2.3.3 Implementation

In this section, we discuss the technical implementation of the Permulin concept. First, we explain the hardware components. Second, we discuss the software implementation.

2.3.3.1 Hardware Realization

The current implementation of Permulin (see Figure 2.23) uses a 52"Philips 3-D display mounted on a table frame (see Figure 2.24). The display can alternatively switch between different full-screen images due to its refresh rate of 120 Hz. At the same time, active shutter glasses that switch between transparency levels at high frequency are wirelessly synchronized with the display. Because of the synchronization between the display and the shutter glasses, each individual pair of shutter glasses can be mapped to an individual, unique output, resulting in each user seeing her *private* view (i.e., unique image) on the entire screen. An increasing number of such glasses (both shutter and polarization filter based) and of compatible 3-D display sets are available. The screen refresh rate defines how many separate views can be offered [Perlin *et al.*, 2000]. Our current implementation offers two views. Displays with high refresh rates and corresponding glasses for more than two personal outputs have already been demonstrated [Wu and Zhai, 2013].

As it is the case for most screens, the display used in Permulin emits linear polarized light, matching the linear filter of the glasses. In consequence, the display would appear black when seen from its short side. We added a diffusion film on top of the screen (Kimoto 100 SXE foil [Kimoto, 2012]), scattering the light and enabling an angle independent (360°) view on both private and group views.

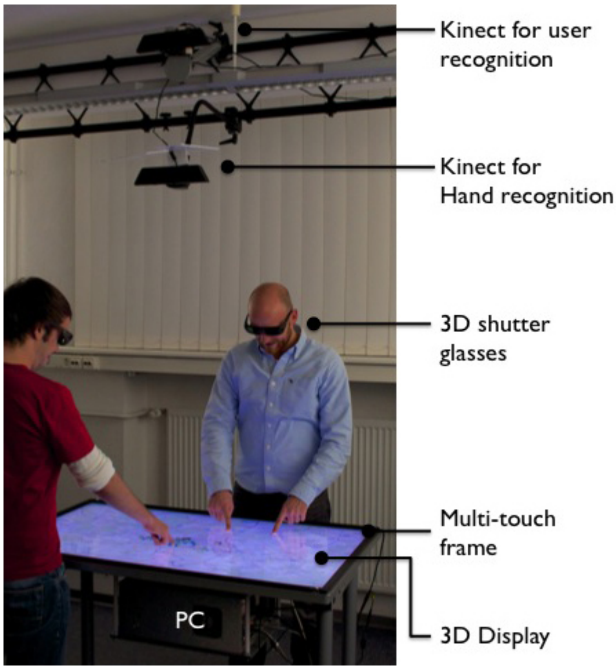


Figure 2.24: Concepual overview

User tracking and hand recognition are based on two Kinect cameras (see Figure 2.24). The higher mounted one caters for user tracking, the lower one detects hands using a contour-based blob tracking approach combined with skin detection. Each newly detected hand is mapped to the nearest user. This mapping is kept as long as the hand is visible to the system, thus leading to personalized hand detection, which is in turn used to assign each touch input to individual users by mapping each touch to the hand contour enclosing it. Touch points are recognized by an infrared multi-touch overlay, supporting up to 32 parallel points.

Readers should note that the entire setting consists of off-the-shelf components and can be easily deployed in offices and meeting rooms.

2.3.3.2 Software Realization

The system architecture of Permulin has several components. A system overview is given in Figure 2.25.

Over the OpenNI-Interface [OpenNI, 2011], which is a 3-D sensing middleware framework, 3-D depth image and color image is accessed. These data are then used as input for the *hand recognition*. The recognized hands are transferred via the OSC-protocol [OSC-protocol, 2014] to the *Permulin Framework*. Hereby, the protocol allows a standard

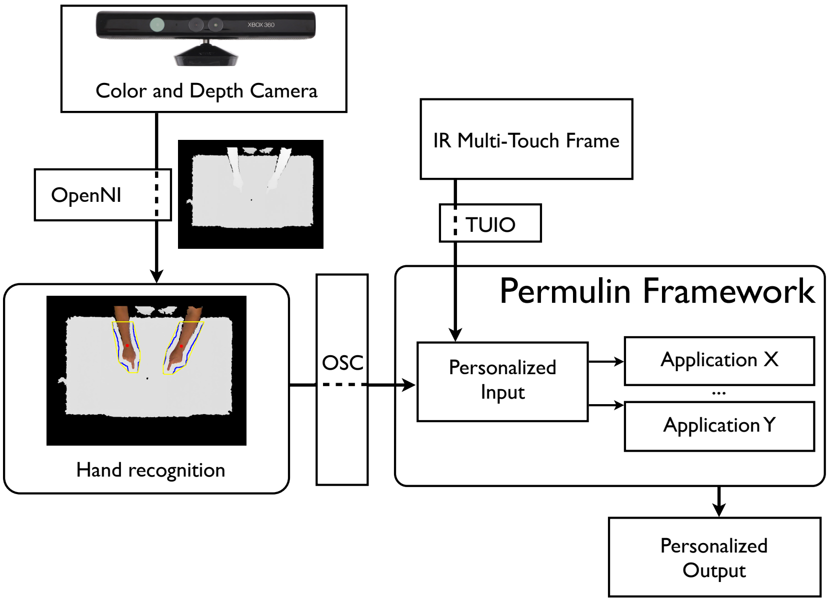


Figure 2.25: Implementation overview of personalized in- and output. The depth image of depth camera is sent to a hand recognition algorithm. Detected hands are then sent to the Permulin framework. Personalized input combines raw touches from the IR multi-touch frame and hand contours to personalize users' input. Personalized input that is mapped to users' application window allows to show personalized output in the 3-D display.

representation and definition of the transferred data. The modular design of the protocol allows that other applications can be loaded and uses this personalized input without re-implementing parts of the code. This protocol combines the data received from hand recognition algorithm with the multi-touch points received from the multi-touch frame. Multi-touch points inside the hand contours are mapped to the corresponding hand and further to the corresponding user, which generates personalized input. The personalized input can then be used to allow personalized output that is also implemented in the framework. Hereby, the personalized input is mapped to the corresponding application window of a user to allow manipulation only on her own view. Details of both personalized in- and output are explained in further detail in the following two sections.

Personalized Input

In order to implement a personalized touch input different sensors are interconnected and combined. Two Kinect cameras located above the horizontal screen are used to combine *user tracking* and *hand recognition* to distinguish which hand belongs to which user. Touch points occurring on the multi-touch frame are mapped to the nearest user's hand. In

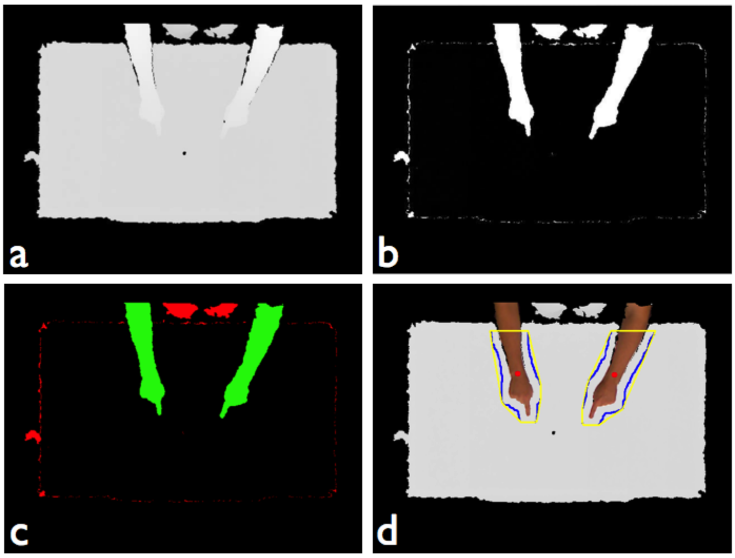


Figure 2.26: Hand recognition step-by-step a-d: (a) raw image of the depth sensing camera, (b) binarized background subtracted depth image, (c) blob and skin detection (green are recognized hand blobs, red are blobs that are not considered due to wrong size or color), and (d) recognized hand position and contour.

the following, we explain in detail how we have implemented user tracking and hand recognition. Furthermore, we compare our Permulin implementation with the related work focusing on personalized input (see subsection 2.2.2).

User Tracking

Depth image from the upper Kinect with a simple blob tracking, recognize where a user is standing.

Hand Recognition

Depth image from the lower Kinect with a computer vision approach recognizes each user's hands (for pseudo code see Algorithm 1). Both outcomes combined map a hand to the corresponding user. Different steps are showcased in Figure 2.26.

In parallel, touch points are detected by the IR multi-touch frame and are transferred to the Permulin Framework via TUIO protocol [3] [Kaltenbrunner *et al.*, 2005]. These two data streams are then combined into a data stream and used for personalized input.

[3]The *Table-Top Tangible User Interfaces*-protocol is mostly used, for example, in Microsoft Pixelsense application to transfer multi-touch events.

Algorithm 1 Algorithm for hand recognition

```
procedure TrackHands(grayImage, colorImage, hands)  ▷ hands serves as input
(hands from last frame) and output (hands from current frame) set.
    applyBackgroundSubtraction(grayImage)
    threshold(grayImage)                              ▷ Apply depth threshold.

    for each blob in blobs do
        checkBlobSize(blob)                                    ▷ see (1)
        checkAmountOfSkinPixels(blob, colorImage)              ▷ see (1)
        checkDensityOfSkinPixels(blob, colorImage)             ▷ see (1)
        checkThicknessOfBlob(blob)                             ▷ see (1)

        img ← and(blob, grayImage)                    ▷ Remove other blobs.
        dilate(img,factor) .       ▷ Amount of expansion (factor) configurable.
        dilatedBlobs ← detectBlob(blob)
        largeBlob ← getLargestBlob(dilatedBlobs)

        contour ← getContour(largeBlob)
        hand ← transformContourIntoPolygon(contour)
        center ← computeGravitationalCenter(largeBlob)

        nearestHand ← findNearestHand(center, hands, range     ▷ Search through
(existing) hands in certain range.
        if nearestHand != null then
            remove nearestHand in hands)
        end if
        add hand in hands
    end for
    return hands
end procedure

(1) Only proceed if within bounds.
```

Comparison of Permulin Implementation with Related Work on Personlized Input

Next, we compare previously introduced techniques for personalized input (subsection 2.2.2) with personalized input realization in Permulin (see Table 2.3).

Permulin provides a reliable hand detection with the accuracy of 94.17% (I3). In comparison to Carpus [Ramakers *et al.*, 2012], this value is lower: accuracy of 97%. Carpus as well as Fiberio, however, requires a high-resolution camera for preciser hand or finger detection. Permulin uses only Kinect depth cameras, allowing to reduce the cost of additional hardware in comparison to the aforementioned camera-based technologies. The mobility is, however, due to the camera approach limited. The system, however, is easy to deploy (I5 partially fulfilled). Because depth cameras are sensitive to sunlight, hand recognition can be effected while operating under sunlight (I3). In contrast to the systems with extra objects, the presented hand recognition operates without instrumentation of the user (I4).

The method allows simultaneous mapping of touch points to multiple users (I1 and I2). This works reliably for up to four hands. The detection rate is, however, reduced to 19 frames per second when four hands interact simultaneously. The reduction does not lead to a noticeable effect while interacting with the system. The frame rate will continue to decline if the system is used by more than two users. One approach to increase the performance would be to process the contour detection in parallel.

A further advantage of the recognition method is that the user can independently choose their position around Permulin. The algorithm is not based on a particular orientation of users' hand to the camera and can thereby be detected in 360°. Because of the independent recognition of the algorithm, the recognition of users' position can be performed by any other algorithm (e.g., proximity sensors). We detect users' position by tracking each user's head with an additional Kinect depth camera, so that all users can freely choose their position (I7). The currently used method for detecting user position does not identify and assign the user so that he can leave and reenter the scene (I6 partially fulfilled).

Requirements for Personalized Input	Simultaneous input I1	Multiple users I2	Robustness I3	No User Instrumenta. I4	Easy Deployment I5	Sustain. User Association I6	Free User Positioning I7	Low cost
DiamondTouch [Dietz and Leigh, 2001]	●	●	●	●	○	◐	○	○
Medusa [Annett *et al.*, 2011]	●	●	◐	●	◐	○	●	○
IR Ring [Roth *et al.*, 2010]	●	●	●	○	◐	◐	●	◐
Smart Floor [Orr and Abowd, 2000]	○	●	●	●	○	●	●	○
Carpus [Ramakers *et al.*, 2012]	●	●	●	●	◐	●	●	○
Fiberio [Holz and Baudisch, 2013]	●	●	●	●	◐	●	●	○
PiVOT [Karnik *et al.*, 2012]	◐	●	◐	○	◐	●	●	○
Permulin	●	●	●	●	◐	◐	●	●

Table 2.3: Comparison of Permulin with presented techniques for personalized input. ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

Personalized Output

In order to realize a personal output, we used a stereoscopic approach with active shutter glasses (see subsubsection 2.2.3.2 for a detailed explanation of related work). In the following, we explain how nowadays commercially available 3-D TV technology allows Permulin to visualize personalized content. Next, we are comparing Permulin with previously presented related work focusing on personalized output.

3-D Technology

While watching 3-D content, active shutter glasses alternately switch between open and closed left or right eye. However, our 3-D TV features a so-called *2 Player Mode*[4]. Hereby, the shutter glasses can automatically be adjusted so that both spectacle glasses of one user are open and the spectacle glasses of the other user are closed. Then both spectacle glasses of one user are opened, and those of the other are closed. This alternately switch, allowing two users as well as two groups of users to see their own two-dimensional personalized content.

Our applications are developed with C# and use Windows Presentation Framework (WPF) library to develop application scenarios. Each user has his own application window that is shrunk to half and occupy only half of the screen. The 3-D display stretches each application to full screen during the two-player mode of the TV so that each user receives a full screen view.

There are several possibilities to transmit 3-D content to the display. For Permulin, we choose the so-called *side-by-side* method. Hereby, the content of both users are positioned side by side next to each other. (see Figure 2.27). The left half of the screen is for user A, and the right half for user B. The *side-by-side* method horizontally stretches each half to full screen and alternatively sends it to the user. In order to have a correct visualization of the content, the application needs to be shrank beforehand with the factor of $\frac{1}{2}$ in the horizontal direction. This process is carried out in the Permulin Framework for all users' content. This process, however, reduces the resolution in the horizontal direction .

Comparison of Permulin Implementation with Related Work on Personlized Output

Next, we compare previously introduced techniques for personalized output (subsection 2.2.3) with personalized output realization in Permulin (see Table 2.4).

Permulin is using a stereoscopic technique with a display refresh rate of 120 Hz, allowing two users or two groups of users to see full-screen independent overlapping views. The number of users is only dependent on the refresh rate of the display and could be increased with an appropriate hardware setting [Kulik *et al.*, 2011; Wu and Zhai, 2013] (O1, O2). For evaluation with two users, our hardware implementation was sufficient.

Permulin is a system that is based on the shutter technique, and users have to wear glasses to see personalized views (O4 not fulfilled). The robustness of personalized output depends, therefore, entirely on how well shutter glasses work together with the display (O3). Tests with Permulin have shown that for the most part, two images could

[4](see page 39, `http://download.p4c.philips.com/files/5/52pfl9606k_02/52pfl9606k_02_dfu_deu.pdf`, last check: 8.11.2013)

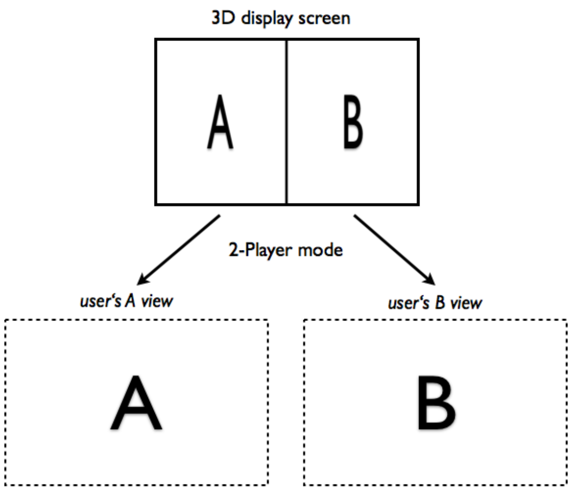


Figure 2.27: Content that is shown to user A and B have to be preprocessed and scaled down with the factor of $\frac{1}{2}$, because the 2-player mode of the 3-D display scales it up to full screen, while presenting the corresponding content to each user.

be achieved; however, this is dependent on the colors used for the user interface. The so-called *Ghosting* effect where the view of the other user is also partly visible, could be observed while using a high proportion of red color. Therefore, a developer should try to use the appropriate color to mitigate this limitation. These properties are, however, dependent on the display device that is used.

A Kimoto 100 SXE [Kimoto, 2012] was placed above the screen to allow a 360° view on the screen, since the (linearly) polarized light loses its polarization when passing through the diffusion film. Without this film, a view is only possible from the longer side of the TV. The user can choose a free position to see his own personalized view (O6). In addition Permulin uses a technique for personalized output that can be used in a mobile setting, because the needed technology is situated within the display (O5).

2.3.4 Example Applications

Two example applications have been implemented to illustrate and evaluate the interaction and visualization techniques. The first one is a full-screen map application that provides route-planning functionality inspired by Tse *et al.* [2004]. The second is a photo sorting application that enables users to co-create a photo collage. Both example applications constitute two highly relevant interface themes: interaction with (i) spatially fixed data and (ii) free-floating interface elements. Both interfaces are illustrated in Figure 2.28.

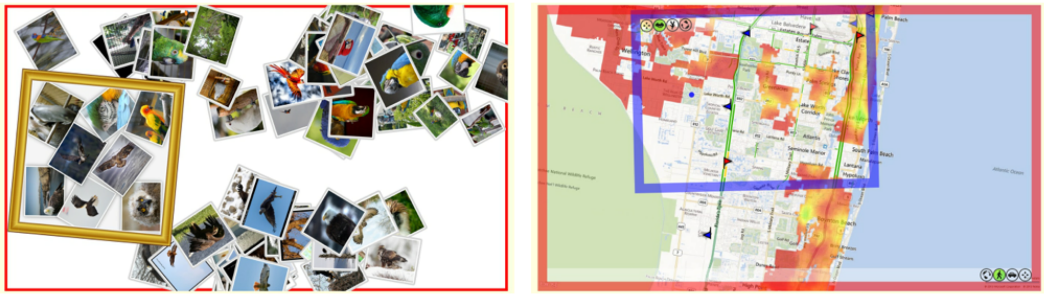


Figure 2.28: Example applications

2.3.4.1 Map Application

The map application displays a full-screen interactive map in the group view that can be explored using conventional pan and zoom multi-touch gestures. The application provides two exemplary visual filters that can be overlaid over the map: a road traffic filter and also a Walkscore filter,[5] to assess the walkability of a neighborhood. The filters are visualized as resizable lenses on the map. A user can place a flag onto the map, indicating the starting position of a route, by tapping and holding. Placing further flags onto the map will create a route that connects all flags in a row.

[5]see http://walkscore.com

	Scaleable Simul. Output	Overlapping	Robustness	User Instrumenta.	Mobility	Free Positioning
Requirements for Personalized Output	O1	O2	O3	O4	O5	O6
Stereoscopic	<6	●	●	○	◐	●
Autostereoscopic	<4	◐	◐	●	◐	◐
Permulin	n [Wu and Zhai, 2013]	●	●	○	●	●

Table 2.4: Comparison of Permulin with presented techniques for personalized output. ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

In case collaborators divide the group view, the maps in the private views are oriented towards the respective users and, together with both filters and flags, can be manipulated individually.

2.3.4.2 Photo Sorting

The photo sorting application visualizes a set of pictures as stackable elements laid out on the tabletop. They can overlap and can be individually manipulated through conventional multi-touch gestures to move, rotate, and scale them. An empty frame, visualized on the group view, serves as a frame for a photo collage. Pictures can be dragged into and removed from the frame.

Collaborators can then either work tightly coupled with all pictures being visible; or they can transition the group view to a private view, where the visibility of the pictures can be toggled through a button on each picture.

2.4 Interaction Concepts

In the following, we present an integrated set of interaction and visualization techniques that effectively support the dynamics of fluid collaboration on multi-view tabletops. All techniques rely on multi-touch gestures, which directly integrate with existing gestures on interactive tabletops. Views and transitions are controlled by multi-touch alone and are fully independent of the user's position and the head and body orientation. All techniques provide support for the main types of contents on interactive tabletops: full-screen contents and free-floating elements, as well as combinations of both.

Presented interaction and visualization techniques (1) provide support both for group work and for individual work, as well as for the transitions in-between; (2) contribute sharing and peeking techniques to support mutual awareness and group coordination during phases of individual work; (3) reduce interference during group work on a group view; and (4) directly integrate with conventional multi-touch input (see Figure 2.29).

In the following, we first present in subsection 2.4.1 interaction concepts for transitioning between individual and group work. While working individually, users still need to synchronize their working state. Second, interaction techniques for that are presented in subsection 2.4.2. Last, concepts to reduce interference while working on a group view are presented in subsection 2.4.3.

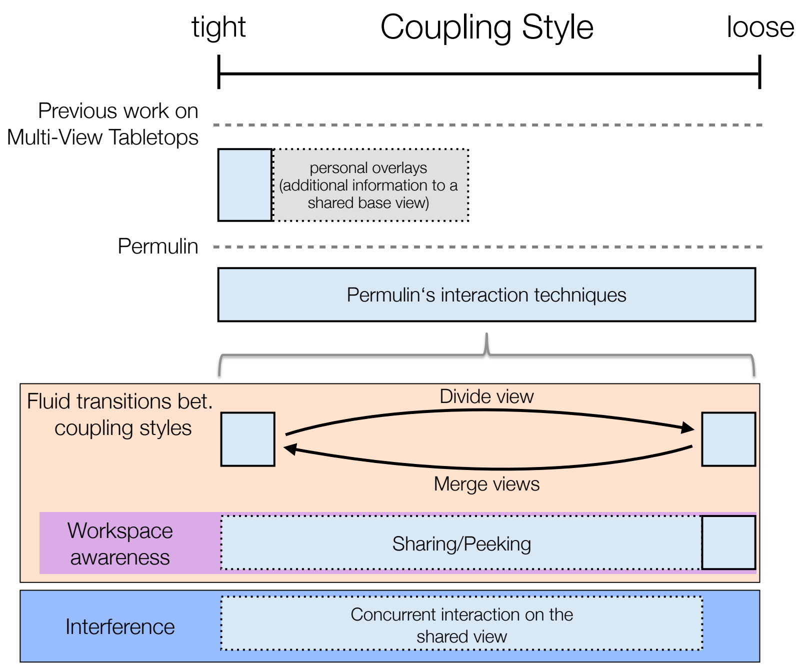


Figure 2.29: This overview presents how Permulin's interaction and visualization techniques allow for fluently transitioning between tight and loose coupling that is key requirement for fluid collaboration.

2.4.1 Fluid Transition between Group and Individual Work

The following two techniques support an easy and seamless transition between a group view that provides common ground during group work, and fully independent views during individual work for each collaborator (R1 as defined in subsection 2.3.1). Permulin provides each collaborator with fully independent, visually overlapping, private full-screen views, whereas previous work on multi-view displays augmented the group view with private contents.

2.4.1.1 Divide Views

This technique transitions the group view to a private full-screen view only for the user performing the gesture. Others remain in the group view. To do so, one of the collaborators performs a gesture that is inspired from grabbing the view: she places her hand flat on the surface and moves it toward her (see Figure 2.30 top). Our implementation for two users creates a private full-screen view for each collaborator, each marked with a user-colored border. If necessary, the view is automatically rotated and oriented to the collaborator. Each private view can be seen and interacted with only by the respective collaborator. To enable several collaborators to interact with their views simultaneously, touch input is personalized and can therefore be mapped to the private view of this user. Initially, the private view is an exact copy of the group view. Subsequently, when collaborators are individually modifying their views, they become different. In consequence, all private views are fully independent of each other and constitute high-resolution workspaces to conduct independent work unobtrusively and loosely coupled.

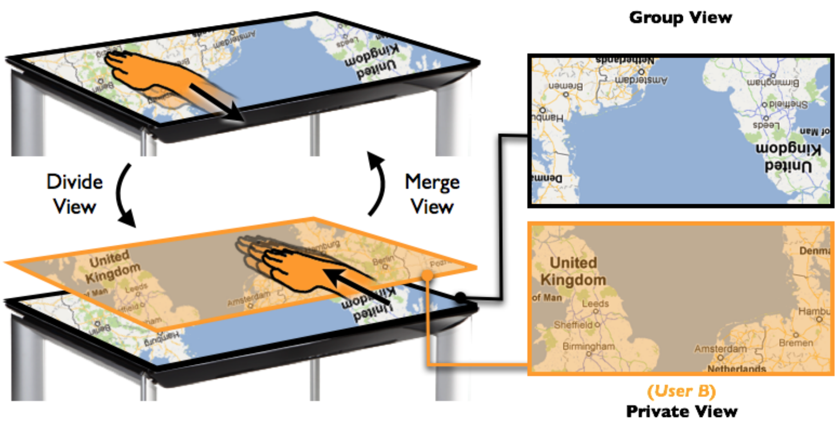


Figure 2.30: Enabling concurrent full-screen collaboration.

2.4.1.2 Merge Views

A private view can be merged back into a common group view to support tightly coupled collaboration. To merge a view, anyone of the collaborators performs a gesture that is inspired from releasing the view. It is similar to the divide gesture introduced above, but performed in the opposite direction (see Figure 2.30 bottom).

The performing user then re-adopts the group view. Hereby, private changes are integrated back to the group view. In case of conflicts (e.g., object changed by multiple users), the object is duplicated and only visualized in the corresponding user's private view with conflicts highlighted. From now on, all manipulations of the corresponding user are again mapped to the group view. Our implementation for two users transitions both users back from their private views to the group view when one of the users performs the gesture.

2.4.2 Awareness and Coordination during Individual Work

Phases of individual work are typically accompanied by moments of tighter coupling, where (portions of) individual workspaces are shared or highlighted, to support mutual awareness and coordination [Tang *et al.*, 2006] (R2 defined in subsection 2.3.1). Previous

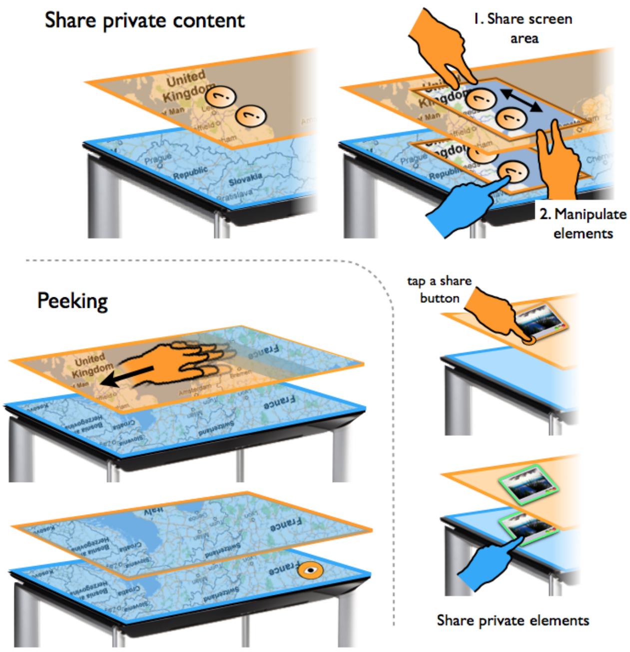


Figure 2.31: Concurrent interaction on group views.

work on multi-view tabletops did not account for sharing of private contents. We contribute two techniques, which support awareness and coordination through sharing and peeking.

2.4.2.1 Quickly and Easily Share Private Content

To share any portion of his or her private view with collaborators, the user performs a pinch gesture with both of her hands simultaneously, i.e. four fingers simultaneously, to avoid conflicts with conventional pinch-to-zoom gestures (see Figure 2.31 top). This frames a shared viewing area, which becomes immediately visible to all collaborators as a window that is overlaid on their view. All collaborators can fully interact with content in this area. The owner can resize the area or maximize the shared view to full screen for sharing her private view in its entirety. Private elements, for example free-floating contents (e.g., overlapping photos), can be shared with other collaborators at any time by tapping a shared toggle button on the top right corner of the respective element and unshared by the same button again.

2.4.2.2 Peek Into a Collaborator's Private View

In the reverse direction, a user can take a peek at another collaborator's private full-screen view (inspired by Bolton *et al.* [2012]) to, for example, quickly assess the work progress of the other collaborator without interfering with her individual work. Figure 2.31 (bottom) illustrates the technique: the three-finger gesture is inspired from temporarily pushing one's own view aside. This reveals the collaborator's view. If more than two users are present, the collaborator has to choose the target user in her private view. Permulin provides awareness thereof to the other collaborator by displaying an eye icon in her private view. A three-finger gesture in any horizontal direction brings the user back to her private view.

2.4.3 Reducing Interference in Group Views

When two or more users would like to interact with different shared elements that overlap in the group view, they are confronted with an access conflict. This becomes particularly problematic when layered content cannot or only hardly be moved, for example, pop-up windows in map applications. Simultaneous interaction on overlapping shared elements can lead to interference. The following techniques try to reduce this interference (R3 defined in subsection 2.3.1).

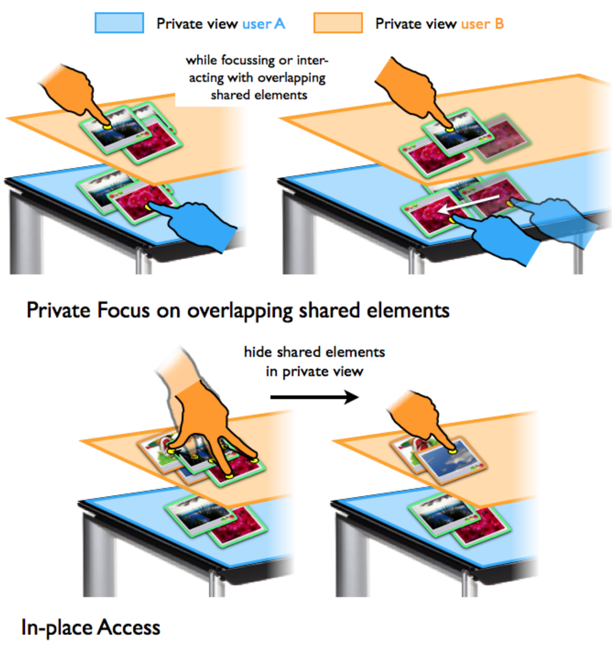


Figure 2.32: Concurrent interaction on the shared view.

2.4.3.1 Private Focus on Overlapping Shared Elements

This technique allows users to concurrently interact with overlapping contents without losing the focus on the user interface element. Figure 2.32 (top right) illustrates this: to enforce a private focus on an element, a user touches and holds the element. The element is then visualized in the foreground in her private view. If multiple users perform this technique on overlapping elements, each of them sees the element they touch in foreground in their respective private view.

2.4.3.2 In-Place Access

When content in the background is occluded by shared layered interface elements in the foreground, e.g. shared by a collaborator, users can hide these elements to reveal content in the background. Figure 2.32 (bottom) illustrates the technique: spreading out three fingers across a pile of foreground elements hides them and reveals underlying elements. The elements are only hidden in the user's private view, not in the group view. This way, collaborators are not disturbed. The reverse action, a three-finger pinch, brings hidden elements back to the fore.

2.4.4 Summary

As mentioned, Permulin allows a fluent switch between shared and private information space. Here, private information space is corresponding to loosely coupled collaboration, and shared information space is for tight coupled collaboration. Presented interaction techniques allow a user to easily switch between loose and tight coupling. In order to summarize the presented interaction techniques, we mapped how this interaction techniques allow a user to fluently transition in the fluid interaction space.

2.5 Evaluation

We conducted a user-centric evaluation to assess the impact of the interaction and visualization techniques on mixed-focus collaboration on multi-view tabletops. The evaluation was a two-step process:

1. A qualitative study was conducted (a) to explore how participants used Permulin in different collaborative coupling styles and what their user experience was, as well as (b) to investigate physical interference's that might occur when users simultaneously perform touch input in overlaid private views.

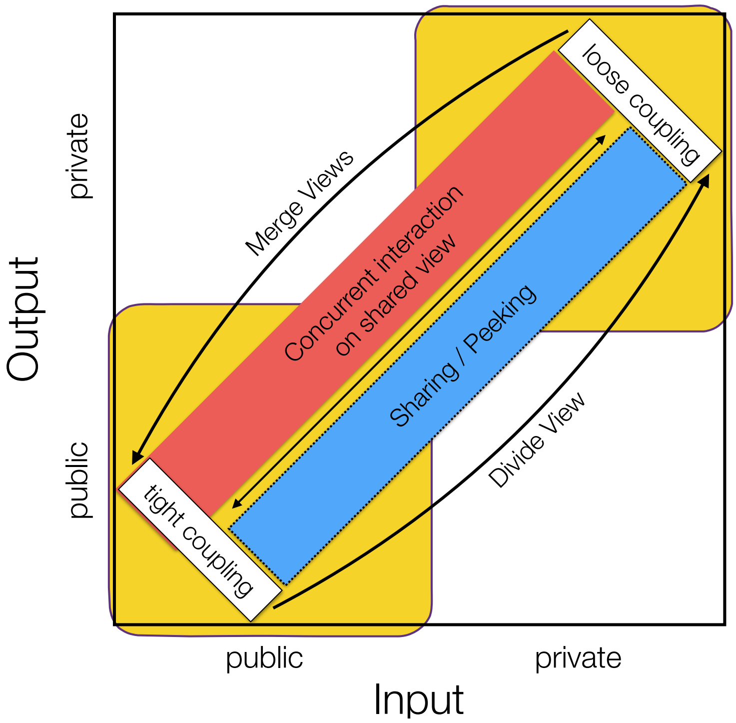


Figure 2.33: This overview presents how the contributed interaction techniques allow a fluent switch between shared and private information space.

2. These results informed a controlled experiment. Permulin was compared with a tabletop system and a splitscreen tabletop regarding (a) collaborators' use of space, (b) their ability to work in parallel, and (c) mutual awareness; all across different coupling styles.

2.5.1 Study 1: Qualitative Exploration

2.5.1.1 Procedure

The participants were asked to collaboratively plan a trip using the map application, inspired by Tang *et al.* [2006]. There were 5 tasks in total. First, participants had to collaboratively search for interesting places in a city of their choice: once without (T1) and once with (T2) the ability to split and merge

Overview – Study Design	
Method:	Qualitative study
Participants:	5 groups of 2 users
Duration:	avg. 2.5h
Data gathering:	Think-aloud protocol, video-taped, interaction logs and semistructured interviews after each task

views. Next, (T3) they started with split views and were asked to coordinate their planning activities from the prior tasks. Afterwards, they had to fulfill a new planning task, starting on the group view (T4). Last, they had to freely plan a city trip, again of their choice (T5).

There were five groups of two volunteer participants each (female, 3; male, 7; mean age, 26 years). Two groups, P1-P2 and P3-P4, consisted of close friends; P5-P6 were friends from work, and two groups, P7-P8 and P9-P10, were strangers. We chose a within-subject design. For each task, participants were given time to familiarize themselves with the system until they felt confident. Each group session lasted about 2.5 hours (think-aloud protocol [Shneiderman and Plaisant, 2004], video-taped, interaction logs and semistructured interviews after each task). After each session, we transcribed the data, selected salient quotes and coded them using an open-coding approach.

2.5.1.2 Results and Discussion

Support of Coupling Styles

All groups used to transition between the group view and the private views when the task setup allowed them to (i.e., in all tasks but T1). Particularly in T2 and T5, they spend long periods in the private views. Throughout the study, participants stressed that the private view helps them focus on individual tasks; as P3 put it: *"I don't have to wait, I can just do my own things [...] and the system helps me to focus on them"*. This is underlined by a strong sense of possession: participants described the surface as *"my territory"* (P5), *"my virtual space"* (P2), and *"my map, and you [P8] have your own map"* (P7).

Despite long periods spent in the private views, participants expressed a strong feeling of cooperation: *"It was always about cooperative work"* (P5, P6) and *"although we worked individually, we still worked together"* (P3, P4). The sharing technique was frequently used to let the other user know about one's own activities, for example, about what they had found on the map. P7 commented: *"It's easy to synchronize different views [...]; it's just there, in front of you."*

The peeking gesture was used by seven users frequently, when the functionality was provided (except T1). Participants particularly appreciated the unobtrusiveness of the technique: *"It does not end my individual work and does not interfere with my collaborator's work"* (P7). Participants further pointed out that peeking allows for quick and easy coordination of their individual workspaces, for example, P5 asked P6 to peek into her view, stating: *"Can you look at my view? I want to show you something."*

However, three participants expressed some uncertainty about what their collaborator was able to see and what not: *"I didn't realize that you could see that [the map in T4]"* (P10).

Physical Interference

We observed that participants frequently interacted in close proximity on the tabletop while they were working on their separate private views. Surprisingly, this did not lead to any notable physical interference, for example, problems created by simultaneous touch input at similar locations. Participants stated that they *"faded out the other participant's fingers"* (P1) and that *"fingers are not problematic, I didn't realize them"* (P9). This finding is further backed by an interesting mismatch between our interaction logs and the participants' perception: The logs show that almost the entire surface had been used for interaction (see Figure 2.34); however, all participants expressed the feeling they had interacted only in their proximity.

2.5.1.3 Summary

We assess the results of the explorative study as very promising. An overview over our results can be found in Table 2.5.

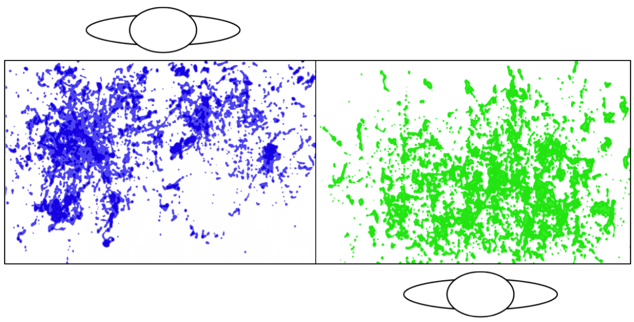


Figure 2.34: Accumulated touch logs aggregated for all participants for T2 and T5 (more intense color represents more touches).

Study Results:
– The overall user experience was that of a personal device during individual work and that of a highly cooperative device during group work.   – The techniques supported users to quickly and easily switched between private and group views (particularly in tasks 2 and 5), allowing them to work in parallel.   – Participants had a strong feeling of collaboration, also when working in the private views.   – Furthermore, participants experienced only little physical interference while using nearly the entire screen for interaction.

Table 2.5: Results of the explorative study.

2.5.2 Study 2: Controlled Experiment

The three major observations derived from the first study provide the basis for a more in-depth investigation of fluid collaboration on multi-view tabletops. In particular, we investigated the following hypotheses:

H1: In co-located mixed-focus collaboration, Permulin provides a larger interaction area than conventional tabletops.

H2: Permulin supports highly parallel work, comparable with a splitscreen tabletop.

H3: Sharing techniques on Permulin to coordinate workspaces are particularly used during mixed and loose coupling. More during mixed than during loose coupling.

H4: A user's awareness over where and what the other collaborator is interacting with and working on

- **H4.1:** does not considerably vary across coupling styles on a multi-touch tabletop with a single view.
- **H4.2:** does considerably vary across coupling styles on Permulin, enabling to transition between high awareness during group work and low awareness during independent work.

2.5.2.1 Setup

Overview – Study Design	
Method:	Quantitative evaluation
Participants:	16 groups of 2 users 8 groups per scenario
Duration:	avg. 2.5h
Data gathering:	Interaction logs, Video recordings, and questionare.
Independent Variables	
Applications:	Photo Collage Route Planing
Device conditions:	Tabletop Split screen Permulin
Coupling styles:	Tight Coupling Mixed Coupling Loose Coupling
Dependent Variables	
Interaction on surface and participant's awareness	

We controlled for three independent variables: the application scenario, the utilized device type, and the coupling style between two collaborators.

As **application scenarios**, we used the two example applications described above. The example applications constitute two highly relevant interface themes: first, the map is a full-screen interface that contains spatially fixed data; moving the data implies moving the map, which is likely to generate interference, and second, the photos are free-floating interface elements that can be arbitrarily moved, rotated and resized on the screen and likely to be stacked.

Three **device conditions** were compared (see Figure 2.35) and were all run on the same hardware prototype. As baselines for comparison, we chose (i) *Tabletop*: a multi-touch tabletop (i.e., with a single view) and (ii) *Split screen*: a tabletop with spatially separated interactive spaces for both users. (iii) *Permulin*: a multi-view tabletop with the techniques and visualizations contributed in this paper.

Moreover, we distinguish between three different **coupling styles**: *Tight Coupling (Tight)*: Working on the same problem; *Mixed Coupling (Mixed)*: Working on the same problem with different starting points or constraints, for example, different pictures or different interests (filters) while planning a route; and *Loose Coupling (Loose)*, working on completely different problems. The dependent variables were (1) interaction on the interactive surface, that is, number, location, and time of touch contact, and (2) a user's awareness over where and what the other collaborator is interacting with and working on.

2.5.2.2 Tasks and Procedure

The study comprised two tasks (see applications in Figure 2.28).

Photo Collage: The participants were asked to create a photo collage using the example application introduced above. At the beginning of the task, the participants were given one or two predefined sets of photos (50 photos each), visualized as a stack. The photo collage was considered finished when participants were satisfied with their results.

Route Planning: The participants had to plan a route using the implemented map application. Each participant had a lens that augmented the map with additional information (traffic and walk score). The task was considered completed, when the participants had found a route.

Table 2.6 gives a detailed overview over the concrete tasks for each coupling style. The coupling style determined the starting situation, i.e. whether participants were in private views or started with a group view on common ground.

We crossed both device type and coupling style for each application scenario. In a pre-study, participants considered the use of a split-screen setup in a tightly coupled collaboration unnecessary and equal to the traditional tabletop setting. Based on this feedback, we removed this condition from the main experiment, resulting in eight subtasks per main task. The order of the tasks was counterbalanced using a balanced Latin square.

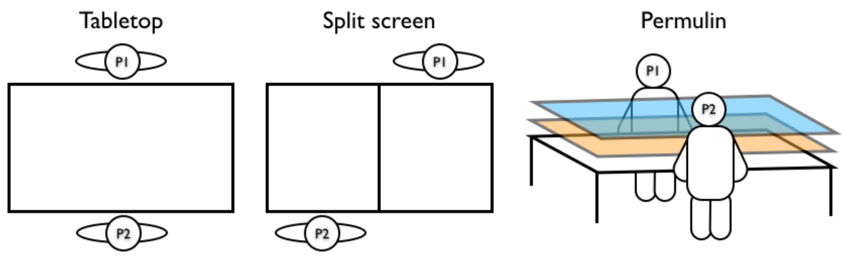


Figure 2.35: Different device types

Coupling Styles		Photo Collage	Route Planning
	Tight	**Data:** We provided a single set of pictures for *both participants*. **Task:** Participants had to design *one photo collage* together.	**Task:** Participants had to plan a *trip together* and find a compromise route between predefined start and end points, while planning to stop twice on the way for sightseeing.
	Mixed	**Data:** We provided a single set of pictures for *both participants*. **Task:** Participants had to design *one photo collage* together.	**Task:** Participants had to plan a *trip together* and find a compromise route between predefined start and end points, while planning to stop twice on the way for sightseeing. Each participant had his *own constraint* that he was asked to follow (constraints were: traffic, walk score).
	Loose	**Data:** We provided a single set of pictures for *both participants*. **Task:** Participants had to design *one photo collage* together.	**Task:** Each user had to plan a *route separately* between predefined start and end points, while planning to stop twice on the way for sightseeing. Both of the routes started or ended in the same area.

Table 2.6: Task description

We chose a within-subject design and recruited 32 participants, each pair of them forming a group (i.e., 16 groups in total). Each of the groups was only assigned to one of the application scenarios (i.e., eight groups per scenario) due to time constraints. During each task, the participants were facing each other (as in Scott *et al.* [2004]) and standing. All interactions were logged and video recorded. After each task, users were asked to fill out a questionnaire. Each group session lasted 2.5 hours in average.

2.5.2.3 Results

Interaction Area

The interaction area was measured as the average percentage of screen space each user was touching, accumulated and normalized over the task durations. It was calculated from the interaction logs. Across all conditions, the personal area was situated in front of each user with most interaction happening in its center and decreasing linearly towards the border of the screen.

The average interaction area was the largest for Permulin (see Table 2.7), for both loose and mixed coupling in the photo task, as well as all coupling styles in the route planning task. A repeated-measures ANOVA showed that the differences are statistically significant (Photo: $F_{(2,14)} = 13.12$, $p < .001$; Map: $F_{(2,14)} = 4.43$, $p < .05$). Bonferroni corrected post hoc tests revealed that the differences between Permulin and Tabletop are statistically significant ($p < .05$;) in both tasks, the difference between the Permulin and Splitscreen condition was only significant in the photo task ($p < .05$;).

Parallel Interaction

	Tabletop	**Splitscreen**	**Permulin**
		Photo Collage	
tight	**45% (SD 14%)**	-	38% (SD 9%)
mixed	34% (SD 8%)	30% (SD 10%)	**40% (SD 11%)**
loose	35% (SD 9%)	33% (SD 12%)	**44% (SD 15%)**
		Route Planning	
tight	15% (SD 11%)	-	**26% (SD 10%)**
mixed	19% (SD 9%)	23% (SD 10%)	**24% (SD 7%)**
loose	25% (SD 12%)	21% (SD 8%)	**29% (SD 9%)**

Table 2.7: Average size of touch areas for both tasks.

We calculated the average percentage of time (see Table 2.8, relative to and normalized over the task duration) where both users touched the screen at the same time, that is, temporally parallel touches within a time frame of 1 s.

In both tasks, the differences between the Tabletop and the Permulin conditions were statistically significant, as shown by a repeated-measures ANOVA with Bonferroni post hoc correction (Photo: $F_{(1,7)} = 21.5, p < .001$; Map: $F_{(1,7)} = 26.4, p < .001$). In case of the photo task, the effect size is small (η^2 = 0.3). However, the large differences between Permulin and Tabletop in the route planning task constitute a large effect size (η^2 = 0.7).

Coordination and Flexible Transitioning on Permulin

We analyzed how participants utilized the techniques for workspace coordination (peeking and sharing), as well as for transitioning between coupling styles (divide and merge) on Permulin.

Peeking and sharing: During loose collaboration, participants peeked in average 2.25 (SD 2.05) times (avg. proportion of time spent peeking 6.53%, SD 9.07%,) and shared their views an average of 4 times (SD 4.81). During mixed collaboration, participants peeked in average 1.25 (SD 1.85) times (avg. time spent peeking 5.58%, SD 8.26%) and shared their views an average of 5.62 times (SD 3.42). As to tight collaboration, participants peeked in average 1.12 (SD 1.64) times (avg. time spent peeking 2.98%, SD 5.42%) and shared their views an average of 7 times (SD 7.09).

Sharing photos was only possible in mixed and loose collaboration, since all photos were shared by default in tight collaboration. Participants shared 14.3 (SD 9.04) photos on average during loose collaboration. However, the photos were 99.2% (SD 0.9%) of the time only visible in their private views on average. For mixed collaboration, participants shared 26.1 (SD 18.58) photos on average, with being 94.55% (SD 5.31%) of the time only visible in their private views, on average.

	Tabletop	**Splitscreen**	**Permulin**
	Photo Collage		
tight	66% (SD 8%)	-	62% (SD 11%)
mixed	70% (SD 15%)	67% (SD 16%)	62% (SD 11%)
loose	78% (SD 13%)	73% (SD 18%)	72% (SD 7%)
	Route Planning		
tight	23% (SD 8%)	-	53% (SD 17%)
mixed	32% (SD 17%)	52% (SD 17%)	40% (SD 20%)
loose	21% (SD 15%)	74% (SD 10%)	66% (SD 19%)

Table 2.8: Average time participants interacted in parallel.

Divide and merge: In loose collaboration, participants spent 100% of the time in divided views. In both mixed and tight collaboration, we identified two dominant collaboration themes: either groups spent most of the time in merged views or in divided views. In the case of mixed coupling, six of eight groups spent 94% (SD 11%) of the time in merged views, whereas the other two groups spent 92% (SD 3%) of the time in divided views. As for tight coupling, three of eight groups spent 100% (SD 0%) in merged views, whereas five of eight spent 73% (SD 13%) in divided views. The amount of time spent in divided views correlates (r = 0.7) with the amount of workspace coordinations (peeking, sharing).

Figure 2.36 illustrates the collaboration of one of the groups using Permulin in tight collaboration. Most of the time was spend in divided views. The collaboration started with a phase of division of labor, then transitioned to a phase of individual work. During this phase, sharing and peeking techniques were used to scaffold workspace awareness. Finally, the participants merged their working states using coordination and transitioning techniques.

Awareness

The awareness was assessed through a questionnaire after each task. The questionnaire consisted of two main parts: the first part (A1) assessed the participant's awareness over where and what the other collaborator was interacting with and working on. The second part (A2) asked the participant to estimate the awareness the other collaborator had over the participant herself. We thus interpret the results as the average amount of awareness cues a device provided to the user. The average results are shown in Table 2.9.

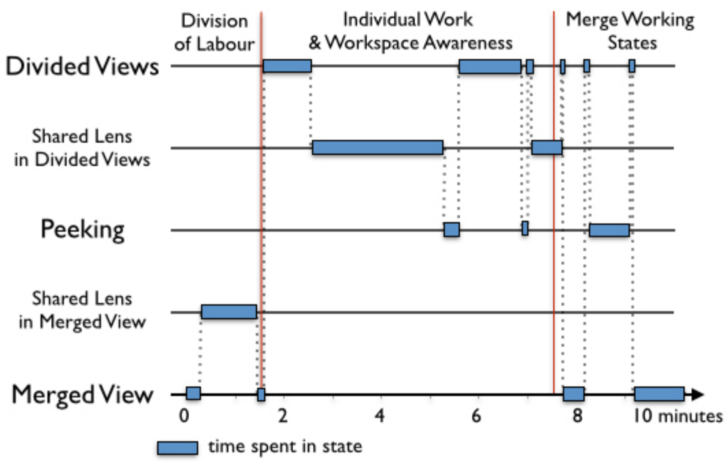


Figure 2.36: Exemplary illustration of a tight collaboration by one of the groups using Permulin.

During loose collaboration, Permulin generated the least awareness cues in both tasks. The difference to the Tabletop and Splitscreen conditions is statistically significant (A1: $F_{(2,8)} = 77.54, p < .001$; A2: $F_{(2,78)} = 77.54, p < .001$) with Bonferroni corrected post hoc tests ($p < .05$ for all differences). In addition, Permulin generated statistically significant less cues during mixed collaboration, while both Tabletop and Splitscreen generated a high amount of awareness cues; as confirmed by a robust repeated-measures ANOVA (A1: $F_{(1.77,69.08)} = 41.24, p < .001$; A2: $F_{(1.77,69.08)} = 41.24, p < .001$) with Bonferroni corrected post hoc tests ($p < .05$ for all differences). Both Permulin and Tabletop generated a high amount of awareness cues during tight collaboration. However, the difference is not significant. As for the Permulin condition, the awareness increased monotonically from loose toward tight coupling, with all differences being statistically significant (A1: $F_{(1.98,77.21)} = 47.61, p < .001$; A2: $F_{(1.99,77.55)} = 43.04, p < .001$; and $p < .05$ for all Bonferroni corrected comparisons).

2.5.2.4 Discussion

When collaborating in a mixed or loosely coupled style, Permulin indeed provides significantly larger personal interaction spaces to the tabletop (H1). This holds even for tight collaboration on a shared full-screen element like a map. In turn, Permulin provides a more open and free interaction space on the very same screen. At the same time, Permulin enables a significantly higher degree of parallel interaction on shared full-screen elements than on the tabletop (H2). The performance is comparable with that of the split-screen tabletop. In the photo task, participants interacted more often in parallel on the tabletop. However, the differences were only little and the effect size was small.

	Tabletop	**Splitscreen**	**Permulin**
	A1: Awareness about where and what the other collaborator was working on (avg. for both tasks)		
tight	**4.37 (SD 0.74)**	-	3.86 (SD 1.32)
mixed	4.37 (SD 0.85)	3.39 (SD 0.97)	2.89 (SD 1.27)
loose	4.11 (SD 1.11)	1.90 (SD 1.11)	1.95 (SD 1.11)
	A2: Estimated awareness of the other collaborator was working on (avg. for both tasks)		
tight	4.05 (SD 0.96)	-	3.61 (SD 1.30)
mixed	3.95 (SD 0.93)	3.41 (SD 0.85)	3.00 (SD 1.18)
loose	3.69 (SD 1.21)	1.87 (SD 0.97)	1.72 (SD 0.90)

Table 2.9: Average ratings awareness questionnaire (1 corresponds to low and 5 to high on a 5-point Likert scale)

Summarized Study Results:
– Permulin indeed provides significantly larger personal interaction spaces to the tabletop (H1). This holds even for tight collaboration on a shared full-screen element like a map.
– Permulin enables a significantly higher degree of parallel interaction on shared full-screen elements than on the tabletop (H2).
– Participants frequently used Permulin's interaction techniques for dividing and merging views, as well as for workspace coordination (H3). This lets us assume that Permulin allows users to easily transition between the coupling styles, for example, when quickly sharing a photo to discuss its importance and then either hiding it again to avoid screen clutter or including it in the collage.
– Permulin provides unique awareness properties: Permulin provides high awareness during group work and unobtrusive work with low awareness during independent work (H4.2).

Table 2.10: Summary of results from the controlled experiment.

Particularly notable is that participants frequently used Permulin's interaction techniques for dividing and merging views, as well as for workspace coordination (H3). This lets us assume that Permulin allows users to easily transition between the coupling styles, for example, when quickly sharing a photo to discuss its importance and then either hiding it again to avoid screen clutter or including it in the collage. The latter is particularly apparent, since participants shared photos frequently during loose collaboration though being only visible in their private views most of the time.

The awareness does not considerably vary across coupling styles on the tabletop (H4.1) and is high for all conditions. The results further show that Permulin provides unique awareness properties: Permulin provides high awareness during group work and unobtrusive work with low awareness during independent work (H4.2). Study results are summarized in Table 2.10.

2.6 Conclusion

In this chapter, we contributed Permulin, an integrated set of interaction and visualization techniques for multi-view tabletops, to support co-located collaboration across a wide variety of collaborative coupling styles. Results from two user studies demonstrate that

- Permulin supports fluid collaboration by allowing the user to *transition fluently between loose and tight collaboration*. The studies show that participants frequently used Permulin's interaction techniques for dividing and merging views as well as sharing content to coordinate workspaces.

- Users utilize Permulin both highly cooperatively but also individually. This is reflected by users occupying significantly larger interaction areas on Permulin than on a tabletop system, as well as performing highly parallel collaboration, particularly on shared full-screen contents.
- Permulin provides unique awareness properties: participants were highly aware of each other and their interactions during tightly coupled collaboration, while being able to unobtrusively perform individual work during loosely coupled collaboration.

3

Spatial Interaction

Contents

In the previous chapter, we presented how users work collaboratively on a novel interactive tabletop called Permulin. We proposed interaction and visualization techniques that allowed for flexibly transitioning between individual and group work by providing each user with full-screen private views. Hereby we studied interaction effects and fluid collaboration on a **single display**.

In the last couple of years, an emergent display technology has evolved that is flexible, thin, and lightweight and has similar form factor and affordances as paper documents. These further called paper-like displays, however, go beyond paper by providing (1) a *high display refresh rate* that supports visualization of high dynamic digital content, and (2) they can be *spatially aware*, meaning knowing about the location of other displays. In this chapter, we present a working concept and simulate these nowadays still unavailable multiple paper-like displays in order to develop interaction and visualization techniques for browsing and viewing dynamic digital content such as videos. These are inspired from the physical world when working in close physical collaboration with paper documents. We investigate how well-known physical interactions with paper documents can be transferred to the area of video navigation for allowing fluid collaboration by spatially interacting with **multiple spatially aware paper-like displays**.

In this chapter, we present CoPaperVideo, a coherent system that allows multiple users to play back and navigate through videos and collections of videos with multiple paper-like displays (see Figure 3.1). It enables users to create an overview of multiple videos as well as structure and organize video contents by leveraging physical arrangements. We contribute a set of interaction techniques for video content that takes advantage of the characteristics of dynamic displays. These techniques go beyond established physical interactions such as arranging and piling of paper. Furthermore, we introduce interaction techniques supporting the management of contents on multiple displays.

Thereby, we advocate a novel collaboration paradigm for videos that consists of using many paper-like displays simultaneously, similar to how we interact and collaborate with printed documents in a group. In this context, we study how interaction and collaboration known from the physical world with documents can be transferred to the world of videos.

The remainder of this chapter is structured as follows. First, we present the scope of this chapter and discuss related work. In section 3.3, we present the design space of multiple spatially aware displays focusing on in- and output. We then describe the CoPaperVideo concept and present supported video activities (see section 3.4). In section 3.5, we present the system design, starting with interactions for spatial arrangements of video, followed by interactions for managing multiple physical displays. Then two iterative implementations of our system are presented (see section 3.6). Lastly, in section 3.7, we present two evaluations that focus on single and multi-user interaction with our system and conclude this chapter.

In summary, this chapter contributes interaction and visualization techniques for collaboratively interacting with multiple videos in physical space. More precisely we contribute the following:

1. An integrated set of novel multi-user interaction and visualization techniques for video navigation on multiple paper-like displays, called CoPaperVideo.
2. A simulation environment and its implementation for simulating and evaluating multiple paper-like displays.
3. Two user studies focusing on single and multi-user interaction with CoPaperVideo addressing the possibility of viewing and structuring multiple videos in parallel. Results from the user studies show that the system effectively supports active video work.

This chapter has been partially published at ACM *Multimedia* [Lissermann *et al.*, 2012b] and ACM SIGCHI Conference on Human Factors in Computing Systems (*CHI*) [Lissermann *et al.*, 2012a].

3.1 Scope of Spatial Interaction

Physical interactions have been widely used specially during physical work with paper documents. People use paper documents for many reasons, particularly for flexible physical navigation, cross-comparison of documents, annotation while reading for adding additional content, and interweaving reading and writing [Sellen and Harper, 2001]. It was also shown that paper documents provide support for *active reading* (i.e., annotating a document while reading for a better understanding) [O'Hara and Sellen, 1997].

We can derive that *sifting and sense-making* of paper-based information is a well-researched field. Research shows that the key is using not only one, but multiple documents or sheets of paper simultaneously, in order to *manipulate and organize information in physical space*. Among others, this has been proven to effectively support comparison, generating an overview and better orientation, also because of the use of physical interactions [Kirsh, 1995; Sellen and Harper, 2001].

In the last couple of years, an **emergent display technology** has evolved that is flexible, thin, and lightweight and has similar form factor and affordances as paper documents [Co and Pashenkov, 2008; Crawford, 2005; Tarun *et al.*, 2013]. These displays, further named *paper-like displays*, however, go beyond paper by providing (1) a *high display refresh rate* that supports visualization of high dynamic digital content, and (2) they can be *spatially aware*, meaning knowing about the location of other displays.

These paper-like displays were our key inspiration for allowing users to work spatially with highly dynamic content such as videos. However, before looking into how paper-based practices can improve working with videos, let's take a look how video work is happening nowadays.

People from a variety of professional backgrounds are confronted daily with large amounts of video footage that they must sift through and make sense of. TV news editors have approximately 30 hours of video material to edit per news agency, such as Reuters,[1] per day. A Hollywood movie director must distill hundreds of hours of footage into a blockbuster movie. Analysts and researchers must make sense of information that is contained within many videos, such as CCTV recordings or recordings of scientific experiments. The YouTube era extends these tasks of sifting and making sense out of many videos to the general population, for hobby and scholarly activities. These examples show that **active video work** with large amounts of video material (as opposed to passive watching of a single video) is a daily routine of many people. Hence, it is obvious that better usability for active video work is a research topic of primary importance. Compared with paper-based practices, today's user interfaces for active video work have three main shortcomings:

(1) Multiple users are mostly restricted to a single screen, so collaborative video browsing is limited in a co-located scenario.

(2) While standard navigation techniques for videos (e.g., play, pause, stop and seeking on a laptop or mobile device) have their obvious benefits, they lack the effectiveness of physical interaction [Kirsh, 1995; Mackay and Pagani, 1994; Sellen and Harper, 2001] for spatially structuring videos.

(3) The traditional "one video at a time" paradigm, where only one video at a time is playing, does not leverage the whole spectrum of human perception. While humans are able to focus only on limited information, they are able to grasp a much higher amount of information in the periphery, which is helpful for getting an overview and structuring.

Inspired from the physical world when working in close physical collaboration with paper documents, we assume that well-known paper-based activities and physical interaction techniques can be effectively transferred to active video work. This could provide support for fluid collaboration by spatially interacting (see Definition 10) with **multiple spatially**

[1] http://www.reuters.com/

aware paper-like displays. This assumption leads to our hypothesis that is the focus of this chapter:

- *How can practices of working with multiple paper documents can be effectively transferred to the domain of video?*
- *How can spatial interactions with paper-like displays support active video work?*

Definition 10 (Spatial Interaction)
In this thesis, I define spatial interaction as a physical interaction with multiple spatially aware paper-like displays. Hereby, the location of all paper-like displays and the user are tracked in 3-D space. Changes in the relative positioning among displays or among user and the displays are translated into spatial input.

Thereby, advocating a paradigm for videos that consists of using many paper-like displays simultaneously, similar to how we lay out multiple printed documents on our desk. In this context, we investigate, how interactions known from physical documents can be transferred to the world of videos. Note that we do not aim to replace existing practices of watching a movie for entertainment purposes. We aim at supporting three activities with videos that are directly related to understanding and making sense out of multiple videos, which involve actively working with multiple videos individually or in a group of people:

1. *Systematic analysis* of videos, for example, for learning or research, similar to active reading of text documents [Adler and Van Doren, 1972]. The main activities to be supported here are: sorting, comparing, and (re)structuring.
2. *Playful video exploration*. Browsing, exploring, and navigating in multiple videos for example, at installations in museums or shops, by providing an intuitive way of interaction. Hereby, playful stands for an intuitive and easy way of browsing and exploring multiple videos.
3. *Lightweight video editing*, where people select and combine video clips, for example, to create a personal excerpt. In contrast to professional video editing tools, we try to provide an lightweight, easy, and engaging way of video editing.

In this chapter, we present CoPaperVideo, a coherent system that allows multiple users to play back and navigate through videos and collections of videos with multiple physical displays (see Figure 3.1). It enables users to create an overview of multiple videos as well as structure and organize video contents by leveraging physical arrangements. We contribute a set of interaction techniques for video content that takes advantage of the characteristics of dynamic paper-like displays. These techniques go beyond established

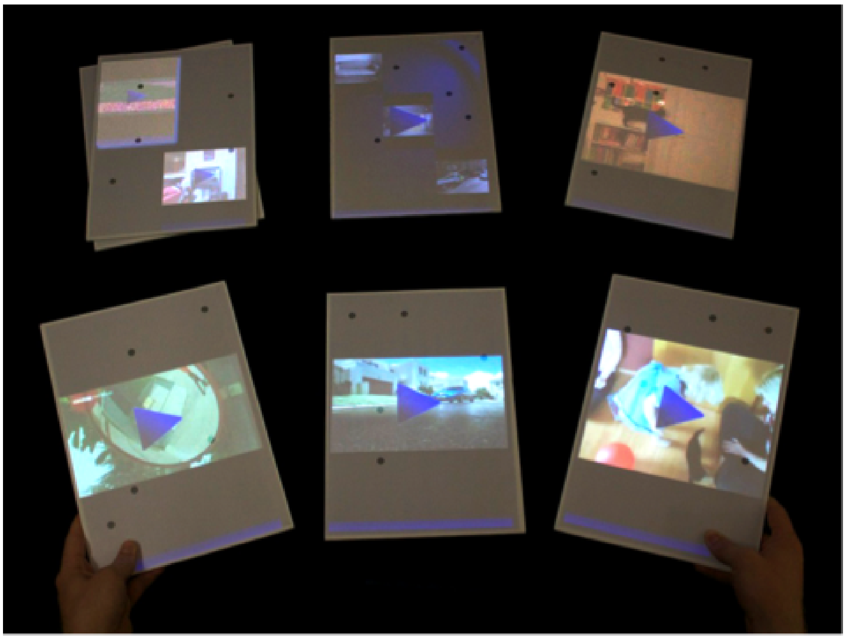


Figure 3.1: CoPaperVideo: A system for collaboratively interacting with multiple videos in physical space.

physical interactions such as arranging and piling of paper. Furthermore, we introduce interaction techniques supporting the management of contents on multiple displays. Finally, we present two evaluations that focus on single and multi-user interaction with our system and conclude this chapter.

In the following, we present related research projects.

3.2 Related Work

In the following, we first place our work in context. Then three related research areas are presented. Finally, we present our design guidelines derived from the related work and compare related work with respect to these guidelines.

In 1993, Fitzmaurice was the first one to present a *single spatially aware tangible display* called Chameleon [Fitzmaurice, 1993]. This display knew his physical context and showed digital information depending on the physical location of the display. For example, dependent on the position of the Chameleon display on a physical map, different additional digital information could be shown on the display. In 1997, Small and Ishii [1997] introduced a spatially aware portable display that allowed users to browse digital

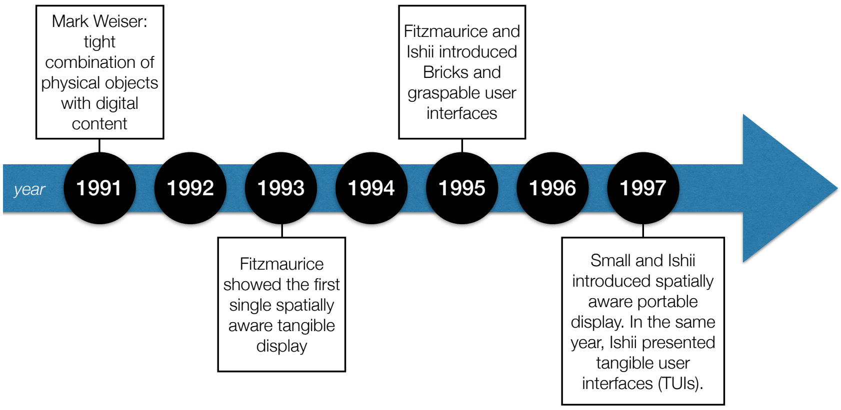


Figure 3.2: The timeline shows research inventions related to a spatial aware display.

information through physical movement. In the same year, based on Weiser's vision, Ishii and Ullmer [1997] introduced tangible user interfaces (TUIs) .

> **Definition 11 (Tangible User Interfaces (TUIs))**
> Tangible User Interfaces (TUIs) will augment the real physical world by coupling digital information to everyday physical objects and environments. [Ishii and Ullmer, 1997]

Tangible user interfaces (see Definition 11) allow users to interact with the digital content by remotely controlling them using physical real-world objects that are visually tracked for spatial input. I think that *multiple* spatially aware displays are a synergy of the ideas presented above. The illustration below contains the chronological order of ideas that, I believe, lead to spatially aware displays. Ishii and Fitzmaurice worked together in 1995 on the incredible idea of TUIs. At that time, they called them graspable user interfaces and were realized with Bricks [Fitzmaurice *et al.*, 1995] that were just used for input. Nowadays, we understand that not only a single spatially aware display can be used for input but also multiple spatially aware displays can provide input and also output both on a single or multiple devices.

In the following, we present previous research related to spatial interaction with videos. First, we will elaborate on why *space management* is of importance for humans. Second, we will present related work on *spatially aware tangible displays* and then we present related work in the area of *video navigation*. Last, we introduce guidelines derived from related work and summarize the related work section by comparing these works based on the guidelines.

3.2.1 Space Management

A large body of empirical work [Kirsh, 1995; Sellen and Harper, 2001; Terrenghi *et al.*, 2007] shows that *physical space management* is important for overview and organization of information and that physical interaction with paper has different qualities physically inspired interaction on touch screens. Previous research has also shown how people structure and use their space while collaborating in a group. Users divide the space in private and group territories [Scott *et al.*, 2004]. Users also separate and partition their space to avoid interference [Tse *et al.*, 2004].

While large tabletop displays would allow for laying out multiple videos in space, interaction on tabletops is inherently limited compared with using multiple physical displays. While tabletop interfaces mimic basic interactions with physical objects, the resulting interaction styles have been shown to be fundamentally different [Terrenghi *et al.*, 2007]. In particular, people make ample use of both hands in physical setups, they mostly restrict interaction to only one hand at a time on tabletops. Physicality also offers a number of advantages such as cues for implicitly assessing the quantity of objects. It is also more difficult for users to arrange objects in a way that can be ergonomically read or viewed on a tabletop compared with with physical displays (this aspect is called micro-mobility [Luff and Heath, 1998] in literature). Finally, tabletops require a static, immobile setup. In contrast, several small physical displays can be used in nomadic setups. While currently available technology does not yet allow us to realize our system without a static setup, nomadic use can be supported in the near future.

Space management and physicality are crucial while structuring and working with information. These works have inspired us to use the rich spatial interactions (e.g., move things in space or piling them) we know from our childhood and bring them to the digital media content such as videos.

3.2.2 Spatially Aware Tangible Displays

When introducing the *first spatial aware display* in 1993, Fitzmaurice stated [Fitzmaurice, 1993]:

> “ [...] Since the information spaces will consist of multimedia data, the display of the palmtop should be able to handle all forms of data including text, graphics, video, and audio. Moreover, the desire to merge the physical and electronic worlds requires that the palmtop computer and display have a *spatial awareness* and understanding of the physical environment along with the ability to *visually mimic these environments and individual objects*. Thus, the combination of a powerful computer capable of understanding and

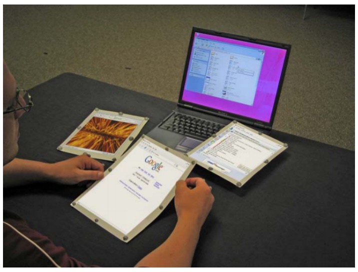
> generating 3-D models coupled with a high-fidelity mobile display will serve to blur the boundaries of the *physical and electronic worlds.* [...]
>
> *George W. Fitzmaurice in 1993* ”

With Fitzmaurice introducing Palmtop, we believe the foundation was laid for *spatial user interaction*. He was the first one who showed a system where the environment knew about the location of the display and the display knew its location inside the physical space. This bidirectional relation is a key requirement for direct physical interaction in 3-D space.

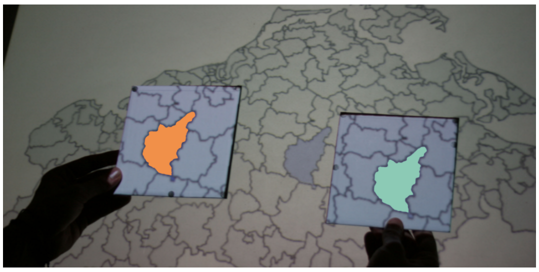
The usage of *multiple physical displays* has been investigated for several purposes. Rekimoto *et al.* [2001] presented a system where multiple transparent tiles were transformed into interactive controls by placing them onto a flat panel display. Siftables [Merrill *et al.*, 2007] demonstrated the technical feasibility of a system of tiny, wirelessly interconnected color displays, introducing multi-display interactions for gaming and educational purposes. Other work presented examples of how several tiny bezel-less screens can be used for interactive board games [Rooke and Vertegaal, 2010] and studied gestures for linking multiple displays [Hinckley, 2003]. A more recent work has investigated how multiple tablets can be combined for cross device interactions based on relative body orientation and position among multiple users [Marquardt *et al.*, 2012]. These works inspired us to provide dynamic visual contents on multiple displays.

Ongoing advances in OLED display technology allow for displaying full-color video on very *lightweight and thin displays*. Interaction with lightweight and thin displays has been a focus of various research projects since the DigitalDesk [Wellner, 1991] introduced the first paper-like display user interfaces, such as a projected virtual calculator. The calculator was top-projected on piece of paper on a conventional table.

Within this stream of research, PaperWindows [Holman *et al.*, 2005] is a very influential work (see Figure 3.3a). It was the first to present a user interface that is distributed over a set of very thin and lightweight paper-like displays. PaperWindows further contributed a set of interaction techniques for basic windowing tasks; however, it did not address interaction with videos. Spindler *et al.* [2009], presented PaperLens a system demonstrating how volumetric or layered data sets can be navigated by the use of paper-like displays. Furthermore, Spindler *et al.* [2010] presented Tangible Views that introduced an interaction vocabulary for multiple paper-like displays that were used above a tabletop to spatially browse 3-D digital content (see Figure 3.3b). Spindler *et al.* improved upon this idea and recently presented a solution for bringing such paper-like displays to the masses [Spindler *et al.*, 2014] by realizing the tracking of such displays with a depth sensing camera.



(a) PaperWindow uses paper-like displays for basic windowing tasks. This figures were taken from Holman *et al.* [2005].



(b) Multiple paper-like displays are used above the tabletop to interact with additional spatial information. These figures were taken from Spindler *et al.* [2010].

Figure 3.3: Influential related works in spatially aware tangible displays.

Our design space (see section 3.3) was inspired by previous works [Holman *et al.*, 2005; Spindler *et al.*, 2010] that have investigated gestures and an interaction vocabulary of paper-like displays. We improve upon these works by addressing paper-based interaction with both, visual and audio content, by introducing a set of interactions for managing multiple displays and by providing first empirical insights into how people use systems with multiple paper-like displays.

A more recent published work has shown that our concepts can be realized with the use of multiple active E-Ink displays: PaperTab [Tarun *et al.*, 2013] shows how spatial input can be leveraged to structure and interact with digital documents on multiple bendable E-Ink displays (see Figure 3.4). The same group also presented FlexKit, allowing to playback video on such displays with 5 frames per second [Holman *et al.*, 2013]. This rate is still too slow to implement and run a mobile version of our concepts and interaction techniques presented in this chapter.

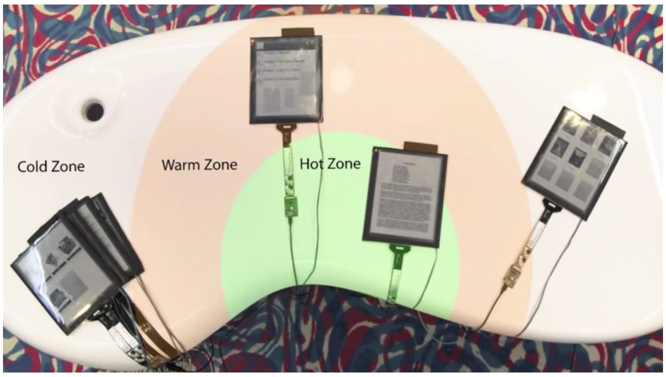
The previous presented related work on spatially aware tangible displays made valuable contributions and inspired our concept. None of these works, however, precisely focused on supporting active video work on multiple paper-like displays that is the focus of our contribution. In the following, we elaborate on related work focusing on video navigation.

3.2.3 Video Navigation

Prior research on *video navigation* investigated interaction *on desktop computers*. In 2007, Glass *et al.* [2007] presented a tool that transcribed, indexed, and summarized lecture recordings that could be accessed and browsed using a Web interface. A project by

[Mertens *et al.*, 2004] presented a desktop interface that allowed a structured navigation in lecture recordings by featuring bookmarking, backtracking, full text search, and footprint. Schoeffmann concentrated on hierarchical video browsing, with 3-D graphics support, for example, tree of playable video segments. Hereby, it provided parallel playback of multiple dynamic thumbnails, while allowing a user to navigate within a video with mouse input [Schoeffmann and del Fabro, 2011]. A profound review of desktop video browsing systems can be found in Schoeffmann *et al.* [2010]. Another project called DRAGON focused on direct and precise navigation in a video by directly clicking and moving the objects in the video stream to navigate in the timeline of the video. This idea was also implemented for mobile use [Karrer *et al.*, 2009].

Video navigation was also explored in the context of *mobile devices*. Hürst *et al.* [2007] explored different types of video timeline sliders for video browsing. He also developed



(a) Active displays are tracked in the physical space.

(b) Selection of content between displays by holding one corner onto the other display.

Figure 3.4: PaperTab is spatially aware active displays. These figures were taken from Tarun *et al.* [2013].

(a) Video Mosaic is the first research project showing tangible video interaction with paper snippets linked to video content. This figure was taken from Mackay and Pagani [1994].

(b) Tangible Video Editor present small active screens that each host a video. Screens can be physically connected and organized in space. Edited video can be played back on a separate screen. This figure was taken from Zigelbaum *et al.* [2007].

Figure 3.5: First research on tangible interaction with videos.

different visualizations for time sliders. A summary of them can be found in his work [Sun and Hurst, 2008]. Huber *et al.* [2010] developed a mobile user interface for interlinked videos. This work quantitatively showed that videos can be browsed more effectively and with less interaction errors using their interface in comparison with a conventional mobile video browser interfaces.

Manske *et al.* [1998] introduced a concept for browsing a video on a large display by visualizing the video as 3-D content tree. Nevertheless, the systems uses only one single screen that is a limitation for multiple users in collaboration. In contrast, CThru [Jiang *et al.*, 2009] combined multimedia content (images, videos, and text) with a storytelling educational video on an interactive tabletop and duplicated the view onto a wall size display. A broader overview can be found in a dissertation of Jochen Huber [Huber, 2012] focusing on interaction with large multimedia information spaces. However, none of these works focused on digital spatial management and spatial cues while handling video content.

Several systems support *tangible interaction with video contents*: Video Mosaic offers a tangible interface for editing video [Mackay and Pagani, 1994]. A snippet of normal paper can be used as a physical token that represents a video. By holding the snippet in front of a camera, the video is played back on a PC screen. A similar approach, using

RFID tokens, is presented in [Sokoler and Edeholt, 2002]. However, these systems do not display the video on the paper snippet, but on a nearby screen. This creates an indirection that is overcome in our work. Video mosaic targets video editing, whereas our focus is on sifting, exploration, and sense-making of video collections. Finally, Tangible Video Editor [Zigelbaum *et al.*, 2007] presented a set of small active screens that can each host a video snippet. By physically arranging displays in a linear sequence, the temporal sequence of clips can be edited. The full video can then be displayed on a computer screen (see Figure 3.5b). This work influenced our approach of physically arranging video displays. In contrast, our work supports large video displays and a wider range of activities and introduces novel tangible interactions.

A more recent published work PaperTab [Tarun *et al.*, 2013], already mentioned in the previous subsection, has shown lightweight, paper-like, active displays that are capable to playback video with 5 FPS. However, these active displays are still either commercially unavailable or too heavy for user studies as is presented in this chapter.

3.2.4 Guidelines and Summary

In the following, we first present guidelines (GL) for spatial interaction with videos on multiple paper-like displays. These guidelines were derived from the previously presented related work. In the following, we first explain the guidelines and then compare most relevant related works to each other.

Space Management

GL1: Physical Space Management The use of physical space provides effective support for overview and organization of information.

Tangible Displays

GL2: Multiple Displays Today's desktop computers and mobile devices usually have only one or two displays. Our system should support a significantly higher number of displays to support physical interactions that are known from the world of paper.

GL3: Spatially Aware To allow for physical interactions that span multiple displays, each display should have knowledge about its relative position in space with respect to its neighbors. Furthermore, each display should have their own 3-D position in space.

GL4: Lightweight These displays should be very thin and lightweight such that they can be easily moved and arranged in physical structures, such as piles.

GL5: Direct Input and Output Each display should also support direct input for navigation purposes as well as sound output.

GL6: High Display Update Rate In order to playback video, the displays should provide color output, high resolution, and high update rate to play back video.

Video Navigation

GL7: Basic Video Controls In order to interact with videos, a basic set of controls such as play, pause, and seeking is needed.

GL8: Simultaneous Playback An important requirement for getting an overview or comparing video content is the ability to playback multiple videos simultaneously.

GL9: Accessing Related Videos Nowadays, related content is often associated with a well-known concept of hyperlinks. For active videos work, it is also important to provide access to this linked video content.

GL10: Video Editing In order to combine and aggregate videos to generate a personal excerpt, specifically while trying to summaries important findings, video editing is a valuable feature.

Fluid Collaboration

GL11: Individual Work The interaction and visualization concepts should support an individual user.

GL12: Group Work The interaction and visualization concepts should support a group of people.

In the following, we will review previously presented related work with respect to the presented guidelines. For details see Table 3.1.

Next, we present the design space for spatially aware displays.

3.3 Design Space of Multiple Spatially Aware Displays

For a systematic design of our interaction techniques, we investigated the design space of how input and output can be performed with multiple spatially aware displays. First, we explain the input for spatially aware displays and then we elaborate on the output.

	Physical Space Management	Multiple Displays	Spatially Aware	Lightweight	Direct Input and Output	High Display Update Rate	Basic Video Controls	Simultaneous Playback	Accessing Related Videos	Video Editing	Individual Work	Group Work
Guidelines	1	2	3	4	5	6	7	8	9	10	11	12
Chameleon [Fitzmaurice, 1993]	○	◐	◐	◐	●	●	○	○	○	○	◐	○
Spatially aware portable display [Small and Ishii, 1997]	○	◐	◐	◐	●	●	○	○	○	○	◐	○
DataTiles [Rekimoto *et al.*, 2001]	◐	●	●	◐	◐	●	○	○	○	○	●	○
Gestures for linking multiple displays [Hinckley, 2003]	◐	●	◐	○	●	●	○	○	○	○	●	◐
Siftables [Merrill *et al.*, 2007]	◐	●	◐	●	◐	◐	○	○	○	○	◐	◐
Hexagonal bezel-less screens [Rooke and Vertegaal, 2010]	◐	●	●	●	●	○	○	○	○	○	◐	◐
PaperWindows [Holman *et al.*, 2005]	●	●	●	●	●	◐	◐	○	○	○	●	◐
PaperLens [Spindler *et al.*, 2009]	◐	●	◐	●	◐	◐	◐	○	○	○	●	◐
Tangible Views [Spindler *et al.*, 2010]	◐	●	◐	●	●	◐	◐	○	○	○	●	◐
PaperTab [Tarun *et al.*, 2013]	●	●	●	●	◐	◐	◐	◐	○	○	●	◐
Our concept we call CoPaperVideo	●	●	●	●	●	●	●	●	●	●	●	●

Table 3.1: Comparison of presented related work for spatial interaction with videos and CoPaperVideo ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

The guidelines are derived as follows: Space Management: GL1. Tangible Displays: GL2, GL3, GL4, GL5 and GL6. Video Navigation: GL7, GL8, GL9 and GL10. Fluid Collaboration: GL11 and GL12.

3.3.1 Input for Spatially Aware Displays

We argue that there are three forms of spatial input with multiple spatially aware displays, reflecting the

- between display(s) and surrounding space,
- between multiple displays, and
- within a given display.

In the following, we describe each input modality in more detail.

1. **Spatial Location Based Interaction**: Moving the display in space is translated into input. Thereby, the absolute position of a display in physical space is captured. This was similarly introduced by Spindler *et al.* in PaperLens [Spindler *et al.*, 2009].

2. **Display Proximity-Based Interaction**: Changes in the relative positioning among two or more displays are translated into input. Interaction primitives include piling of displays and using one display as a pointing device for selecting content on another display.

3. **Display Centric Interaction**: The user can perform physical gestures with one display (also introduced by Spindler *et al.* [2010] in Tangible Views) or with a set of displays, for example, by shaking. Moreover, the user can directly interact with contents on a display using direct touch or pen input.

This allowed us to identify several interaction primitives, grouped along three different basic forms of input (see Figure 3.6). Our system also tracks the users and allow them to control and manipulate spatially located 3-D sound by moving close or away from the displays. This type of input, however, is user centric and is based on proxemics interaction between the user and the displays [Ballendat *et al.*, 2010], and therefore, it is not listed in the design space. Similar interaction techniques have been recently presented by Spindler *et al.* [2012] with 3-D content on paper-like displays.

3.3.2 Output for Spatially Aware Displays

On the output side, our design space distinguishes two modalities:

1. **Visual Output**: Output is directly projected or visualized on the display.

2. **Sound Output**: Another way of output is spatially located sound sources. These can, for example, be generated by spatially located displays with sound output. Another approach to provide spatial sound, is virtually positioning sound sources in 3-D space. This approach is also used in the CoPaperVideo implementation (for more detail, please see subsubsection 4.1.2.3).

Paper-like displays are used to manipulate both output modalities. Spatial interaction concept to navigate and control videos (visual output) with multiple paper-like displays is presented in the next section. Spatial interaction techniques allowing single and multiple users to grasp multiple sound sources in 3-D space (sound output) are presented in chapter 4.

3.4 Supporting Collaborative Spatial Interaction with Videos

In the following, we first present our interaction concepts for spatially interacting with video content on multiple paper-like displays, called CoPaperVideo. Next, we explain our supported scenario for active video work.

3.4.1 Concept

In our concept, called CoPaperVideo, multiple paper-like displays are used collaboratively for video-based activities such as systematic or explorative video navigation as well as lightweight video editing. CoPaperVideo is an integrated set of novel interaction and visualization techniques allowing spatial interaction with video content. In detail,

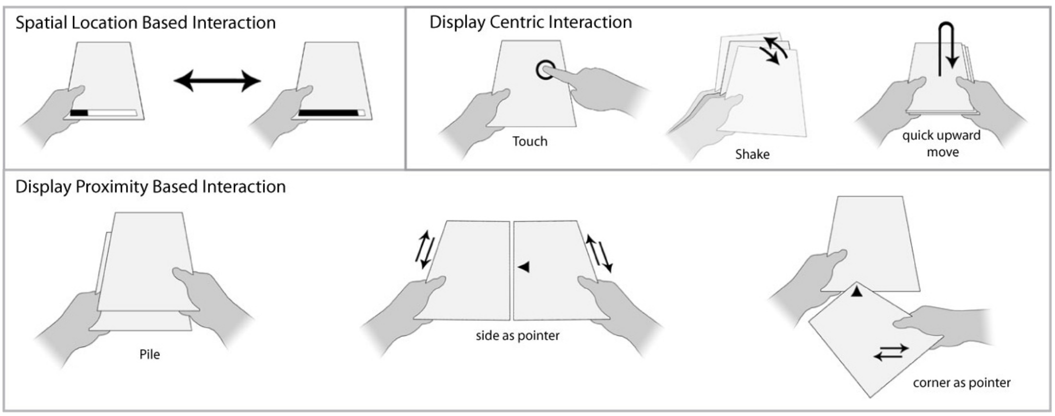


Figure 3.6: Input design space of multiple spatially aware displays.

CoPaperVideo enables users to create an overview of multiple videos as well as structure and organize video contents by leveraging physical arrangements.

To technically realize displays that can be manipulated like paper, CoPaperVideo makes use of a projection-tracking setup. Passive cardboards are tracked in real time. Visual content is automatically projected onto them, correcting for perspective distortions. We envision such systems to be installed at offices, schools, libraries, museums, and stores. Moreover, given the rapid advances in mobile devices, future tablet devices are very likely to be much thinner and more lightweight than nowadays [Tarun *et al.*, 2013]. This will eventually render projection-tracking obsolete and also allow for mobile use cases of our system (see Figure 3.7).

A detailed comparison of related work and CoPaperVideo in respect to previously introduced guidelines can be found in Table 3.1. From the table, it is visible that paper-like displays have been introduced by previous works [Holman *et al.*, 2005; Spindler *et al.*, 2009, 2010, 2014]. By the use of paper-like displays, CoPaperVideo supports the first five guidelines (GL1-GL6). These guidelines are highly important for spatial interaction. CoPaperVideo provides interaction and visualization techniques for video navigation (GL7-GL10) on multiple paper-like displays while supporting individual and group work (G11 and G12).

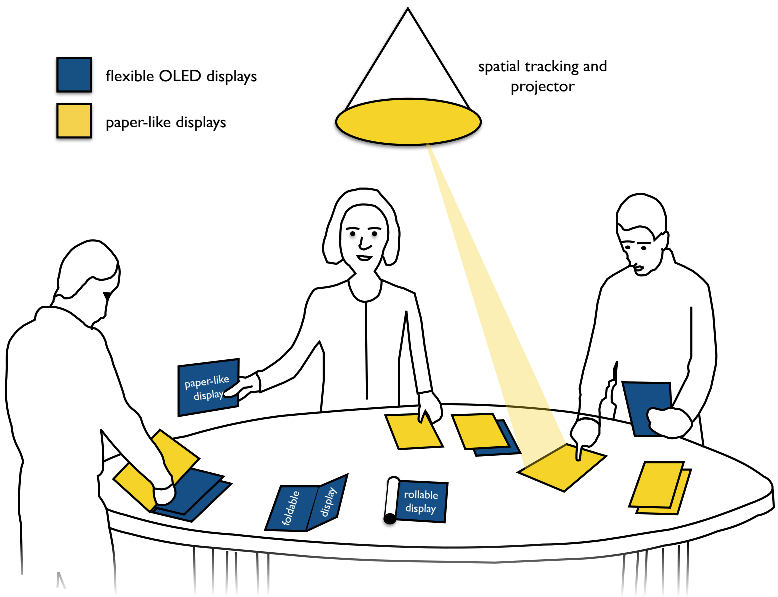


Figure 3.7: **Spatial interaction with spatially aware displays. (Yellow) Lightweight paper-like displays that are tracked in space and content are projected onto the display. (Blue) Flexible OLED displays that feature high resolution and are also spatially aware, used together with paper-like displays.**

3.4.2 Supported Video-Based Activities

With CoPaperVideo, we aim at supporting three main scenarios of active video work, which benefit from multiple video displays:

1. *Systematic Video Analysis*: Active reading [Adler and Van Doren, 1972] is a well-studied domain. Active reading involves intensively engaging with documents, for instance, by following references, annotating, and comparing documents. People often work with multiple documents simultaneously and effectively arrange them in physical space to support their reading. Analogously, we propose *systematic video analysis* as a way of actively working with video material: people explore a set of videos, prioritize the content, study related content, and compare and (re)structure the content. As outlined in the introduction, these activities are of crucial importance in a wide range of professions. Different areas such as film (post)production by novice/professional users or analysts of a huge amount of multiperspective camera recordings (e.g., from a catastrophe) are in need of prioritizing, comparing and (re)structuring video snippets. Based on the mentioned needs of actively working with videos, we are convinced that in the case of videos, the use of space provides effective support for such highly creative and dynamic activities.

2. *Playful Video Exploration*: Multiple displays can be beneficial for exploring collections of videos, for instance, at installations in places like stores, museums, or exhibition booths. We envision videos spread on booths or tables, where visitors can stop by and playfully explore new topics or products individually or collaboratively. The focus of such systems is not only on functionality but also on high user experience as well as ease of access. Furthermore, it enables users to serendipitously discover content. The system should be intuitively usable to facilitate a playful exploration and a positive experience to people of all ages and professional backgrounds.

3. *Lightweight Video Editing*: Simple video editing is a common part of using videos as a consumer. For instance, people trim video snippets or they order and align several snippets in a personal excerpt. CoPaperVideo supports such simple editing tasks. They are conceptually similar to highlighting or excerpting passages in active reading of text documents, which serves for better understanding and condensing the contents. This is opposed to advanced video editing that focuses on production of videos, which includes specific functions like time-stamping or synchronizing footage from multiple cameras over time in order to sort the recordings, annotating and augmenting video snippets or position in a video with additional information.

All of these scenarios have a set of functionality in common: users require functionality for quickly getting an overview of a single or multiple videos, for prioritizing content, finding related content, comparing content, and (re)structuring content. The system should support quick temporal navigation, cross-video use for overview, comparison and linking, as well as flexible means for prioritizing, grouping, and structuring.

3.5 Interaction Techniques with Videos on Multiple Paper-like Displays

Inspired by the physical world and interaction with paper, in the following, we present the interaction techniques for spatial interaction with videos on multiple paper-like displays. We first present interaction with videos by using physical 3-D space, for example, seeing though a pile or accessing related videos. Second, we introduce interaction techniques for managing digital content between multiple displays, for example, combine and distribute content.

3.5.1 Interaction with Videos in Physical Space

In this section, we present interaction techniques that support a set of basic activities for individual videos and collections of videos. These techniques leverage the manipulation and arrangement of one or several displays in physical space.

3.5.1.1 Temporal Navigation

Temporal navigation within a video is one of the most basic functionalities. It is required to get an overview of the video as well as for quickly accessing specific passages. Similar to existing user interfaces for desktop computers and mobile devices, our design allows users to start, pause, and skim a video by directly interacting with widgets on the display using a stylus (see Figure 3.8a and Figure 3.8b). In contrast to most existing interfaces, it is possible to play multiple videos simultaneously.

Space is a strong cue for encoding information, for instance sequences [Kirsh, 1995]. This motivated us to design a technique in which the physical workspace encodes temporal positions. The timeline of the video is virtually spread out in physical space, extending from left to right within the user's arm reach (see Figure 3.8c). Each spatial position is mapped to a temporal position within the video. By moving a display through space (simultaneously moving several displays is also possible), the user navigates through the video. This technique allows for quickly skimming as well as for jumping back and forth

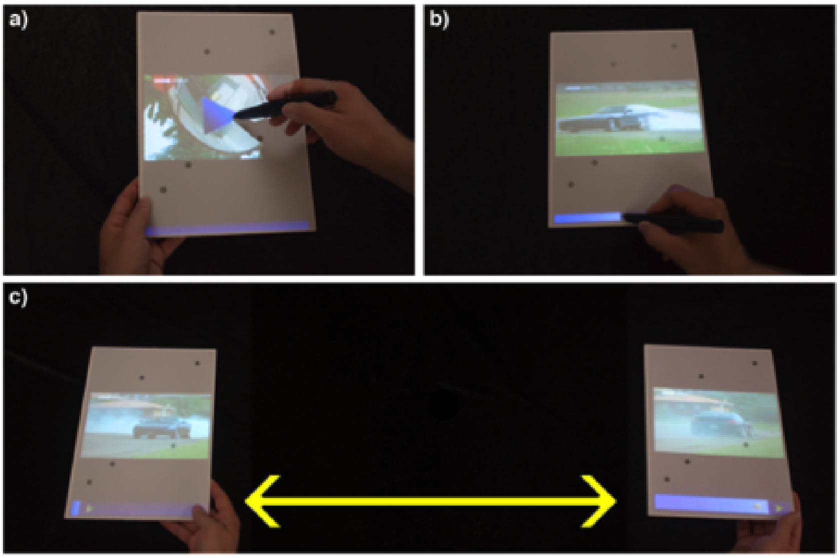


Figure 3.8: a) Playback, b) skimming, c) temporal navigation in space

between several passages of a video. To avoid interfering with free arrangements of multiple displays (see next subsection), temporal navigation is activated when the user lifts the display up to a higher level above the table surface, the "temporal layer."

3.5.1.2 Arranging

Similar to arranging objects in the real world, the thin and lightweight displays can be freely arranged on the table surface. To state only a couple examples, two videos can be compared by placing them side-by-side while multiple videos can be ordered in a spatially encoded sequence. Videos can be prioritized by placing them closer or more distant to the user. Such arrangements enable powerful ways of organizing information in space [Kirsh, 1995].

3.5.1.3 See-Through Pile

A large body of research shows the relevance of piling for managing information [Malone, 1983; Mander *et al.*, 1992]. Users can place multiple displays on top of one another to form a pile of videos. Our pile is more advanced in comparison with piles of ordinary physical objects. Since the system is aware of which displays are occluded, the content of the entire pile is visualized on the topmost display, resulting in an "x-ray style" view (see Figure 3.9). All of the content on the topmost display is fully interactive. So the user can view, play, or skim any video in the pile easily. Piling or unpiling does not interrupt playback; the video continues to play inside or outside a pile.

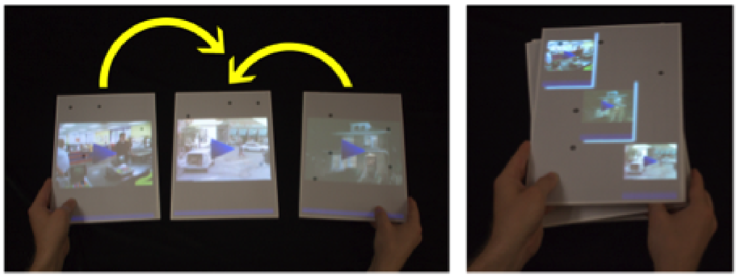


Figure 3.9: Physical piling of videos. The topmost display allows for interacting with all videos.

3.5.1.4 Accessing Related Videos

Many videos are organized in collections, in which they are linked to related videos. This is the case for influential video platforms such as YouTube, iTunes U [Apple, 2007], and OpenCourseWare [MIT, 2002]. We present a spatial technique for navigating video relations using multiple displays. By bringing an empty display near to a display with a video (see Figure 3.10), the mode for selecting related videos is entered. A list of related videos is visualized on the previously empty display. By moving the empty display or the video display up or down, a related video can be selected from the list. While doing so, a preview of the currently selected video is shown on the empty display. By slightly moving one display away from the other, the list shows categories or groups of videos instead of individual videos, allowing for a selection at a higher level. By moving one display apart, the related video (or group of videos) is eventually selected and displayed on the previously empty display (see Figure 3.10).

There are two main advantages in this gesture while working with multiple displays. First, the original video is not replaced by the related one, as in most current solutions, but remains visible. Hence, a spatial overview can be easily generated by leaving a trace of "where we came from." Second, multiple related videos can be opened by using multiple displays. These videos can then be spatially arranged and also viewed in parallel, if desired.

3.5.1.5 Linking Videos

The user can create his own hyperlinks between any two videos. This is done by taking two displays with different videos and bumping them against each other (see Figure 3.11). From now on, the videos appear in the respective lists of linked videos.

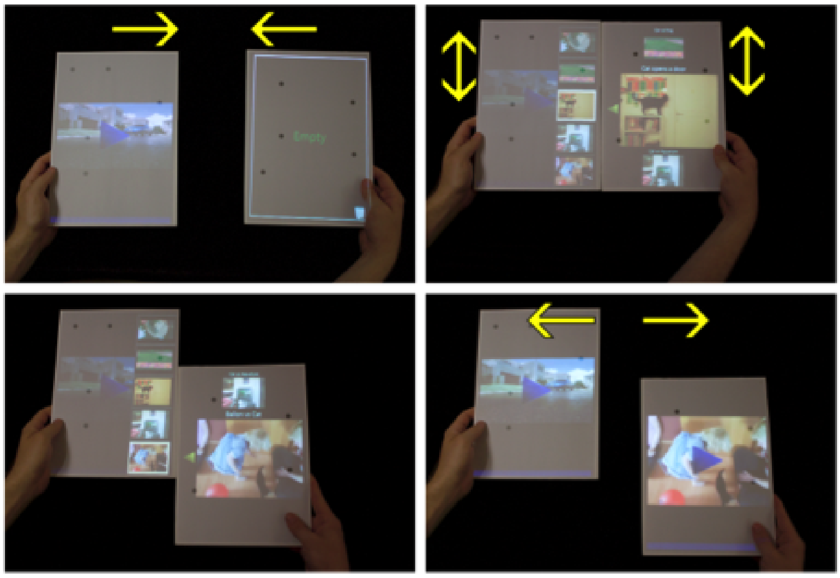


Figure 3.10: Accessing related videos by bringing displays side by side, selecting a related video, and moving the displays apart.

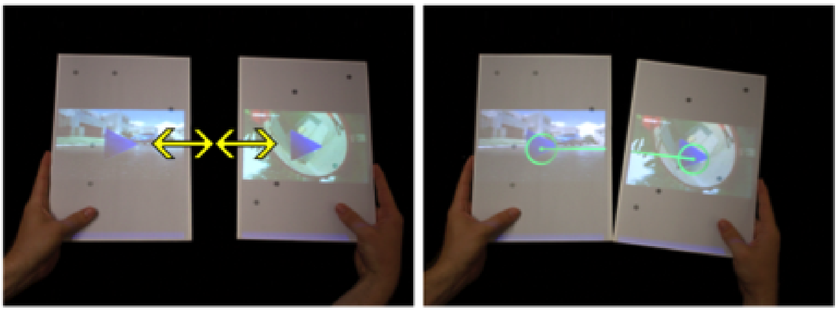


Figure 3.11: Linking two videos by bumping them against each other.

3.5.1.6 Lightweight Video Editing

When people actively work with text documents, they highlight passages that are of high interest, write excerpts, and create text collages by copying and pasting relevant passages into a new document. In contrast, video documents are usually consumed as-is, without personalizing them. We propose a lightweight interaction technique for cutting videos. We do not aim for professional video editing, but for providing a simple interaction technique. This can be used for focusing on specific passages of a video and for composing a "video excerpt."

For cutting out a section of an existing video, an additional empty display is needed. By placing one corner of the empty display onto the timeline of the video, the corners are used as a video cutting tool. The start and end positions of the cut are selected with the

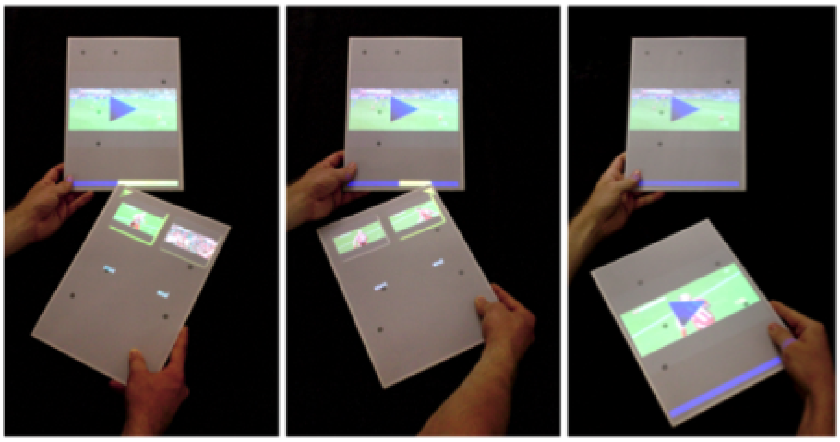


Figure 3.12: Playful video cutting with multiple displays.

upper-left and upper-right corners, respectively. While selecting, the start and end frames are visualized on the previously empty display and the entire passage is highlighted in the video timeline (see Figure 3.12). By moving the display apart, the cut is executed and the newly created video snippet is made available on this display. From now on, the user can interact with this snippet as with any ordinary video. The next section discusses how several physical video snippets can be combined to one video.

3.5.2 Managing Multiple Displays

Paper documents have a static mapping between contents and the physical carrier medium. One page of content is permanently bound to one page of the carrier medium; this is a one-to-one mapping. Most computing devices have only one screen. Here, in contrast to paper documents, the content is dynamic. Potentially, an infinite number of content can be displayed on one physical screen: here, we have a many-to-one mapping. Both types of content mappings are well understood.

Given the assumption that future displays will be low priced and lightweight, we imagine that users will have a number of displays that combined are much smaller than the number of sheets of paper that we typically use with printed documents today. Hence, we are not limited to tight one-to-one mapping of content to displays. This would also not be desirable, as it would limit the display's capability of dynamically changing its content.

Definition 12 (Content Virtualization)
Digital content is virtualized when it is temporarily disassociated from a physical display.

Definition 13 (Content Materialization)
Digital content is materialized when it is bound to a physical display.

Therefore, systems that offer many paper-like displays have a many-to-many mapping. Such systems mimic the physical interactions of paper, however, with a smaller number of carrier media. Previous work has shown how contents can be easily transferred from one display to another [Holman *et al.*, 2005; Rekimoto, 1997]. However, it is not clear

1. how the handling and association of content on many displays should work.
2. how content can be temporarily disassociated from displays to generate free carriers for displaying additional contents.
3. How such "virtualized" (see Definition 12) contents can be "materialized" (see Definition 13) again and bound to physical carriers.

We present interaction techniques that allow the user to combine contents onto one single display, distribute content over multiple displays, and clear and restore content.

3.5.2.1 Combining and Distributing Content

Each video is so far bound to a physical display. By combining these videos to a set of videos on a physical display, other physical displays can be freed from contents. Thereafter, they can serve as physical carriers for additional videos. To combine one or several videos, the user creates a pile out of the respective displays. Quickly moving the entire pile upward combines all videos into the topmost display. The remaining displays inside the pile become empty. The metaphor of this interaction is to push all videos up, which are caught by the topmost display.

The reverse direction, distributing videos from a video collection, is done by placing one or several empty displays underneath a display containing a video collection. By quickly moving the pile downward, the videos from the collection are distributed onto the empty displays in the pile (see Figure 3.13).

3.5.3 Clearing and Restoring Content

Contents can also be virtualized by clearing a display. Clearing is performed by shaking the display, as if one shook contents off. Cleared contents are available in the recycle bin,

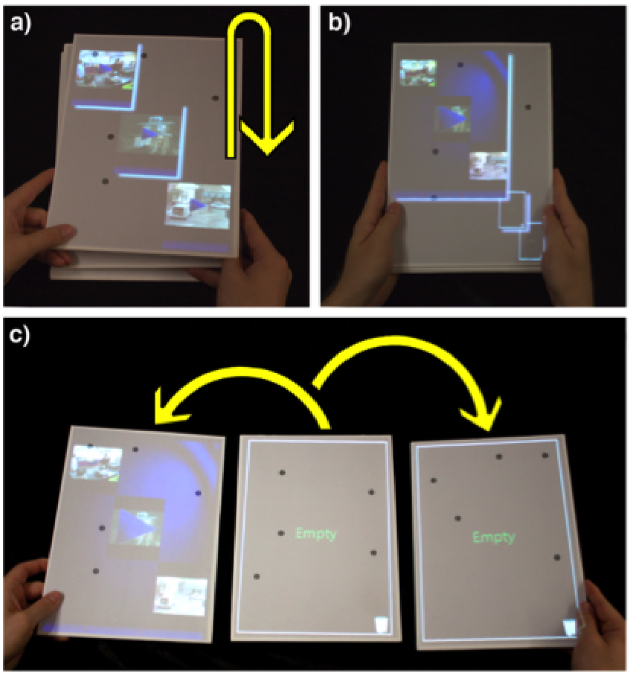


Figure 3.13: Virtualizing a physical pile (a) onto one single display (b). This clears the remaining displays (c).

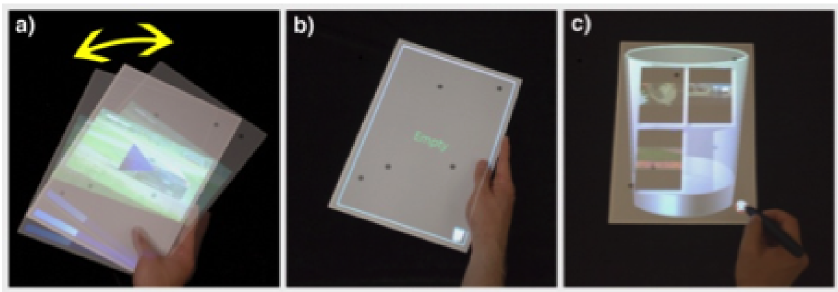


Figure 3.14: Clear content by shaking the display (a and b). Restore content from the recycle bin (c).

which can be accessed by tapping on an icon that is available on all empty displays (see Figure 3.14).

3.5.4 Summary

Table 3.2 summarize presented interaction techniques with videos on multiple paper-like displays. In the following, we explain how we have technically realized our system.

Name	**Purpose**	**Description**
Interaction with Videos on Physical Space		
Temporal Navigation	Start, pause, and skim a video	User can have basic video control with widgets each display. By moving a display through space, a user can navigate through the video.
Arranging	Physical structuring, ordering, and prioritizing	Similar to arranging objects in the real world, the paper-like displays can be freely arranged.
See-through Pile	Piling of videos	Users can pile videos with multiple displays on top of one another. Since the system is aware of which displays are occluded, the content of the entire pile is visualized on the topmost display.
Accessing Related Videos	Physical access of related videos	A spatial technique for navigating related video collections using multiple displays.
Linking Videos	Creating hyperlinks between any two videos.	Two displays with different videos are bumped against each other.
Lightweight Video Editing	Cutting out a section of an existing video	Corners are used as a video cutting tool. The start and end positions of the cut are selected with the upper-left and upper-right corners, respectively. By moving the displays apart, the cut is executed.
Managing Multiple Displays		
Combining Content	Combine contents onto one single display	Quickly moving the entire pile upward combines all videos into the topmost display. The remaining displays inside the pile become empty.
Distributing Content	Distribute content over multiple displays	Placing one or several empty displays underneath a display containing a video collection. By quickly moving the pile downward, the videos from the collection are distributed onto the empty displays in the pile.
Clearing and Restoring Content		
Clearing Content	Clearing content on the display	Clearing is performed by shaking the display.
Restoring Content	Restoring content on the display	Cleared contents are available in the recycle bin, which can be accessed by tapping on an icon that is available on all empty displays.

Table 3.2: Summary of the interaction techniques included in the CoPaper-Video system.

3.6 Implementation

In the following ,we explain our technical realization of the system. First, we explain our initial implementation that focused on a single-user scenario. Last, we explain the enhanced implementation that provided a collaborative use of our system.

3.6.1 Single-User Prototype

3.6.1.1 Tracking and Output on the Display

Our first prototype system realizes paper-like displays by tracking passive cardboards in real time and projecting contents onto them. An overview of the system is shown in Figure 3.15. Our system consists of an optical tracking system (called Optitrack [OptiTrack, 2008]) with six infrared cameras, two full HD projectors mounted on the ceiling, and a set of cardboards, each augmented with infrared retro reflective markers. The high-resolution projection frustum measures approximately $200 \times 120 \times 40$ cm^3. The information that is provided by the tracking system (position, orientation of the cardboards) is used to warp the projected images onto the cardboards in realtime.[2] In our software toolkit, we simulate the environment by constructing a Direct3D world model. In an initial calibration step, the two Direct3D cameras are set to the positions and orientations of the two projectors, thus the camera "sees" the multiple cardboards and renders their contents from the correct perspective. The projectors display the camera views, which are generated by Direct3D, while the world model is continuously updated by the tracker data. For recognizing the different gestures, we implemented a gesture recognizer that analyzes positional information of each of the displays.

The application is implemented in C# and WPF. Each display owns its own WPF window that is screen captured and rendered onto the display on demand. With this implementation, our prototype supports seven different cardboard displays.

3.6.1.2 Input on the Display

Stylus input is realized using an Anoto Digital Pen ADP-301 and the Letras software framework [Heinrichs *et al.*, 2010]. Each cardboard is therefore augmented with the Anoto pattern.

[2] For more detail, please refer to this bachelor thesis [Riemann, 2010]. This functionality has been implemented by a student from our lab.

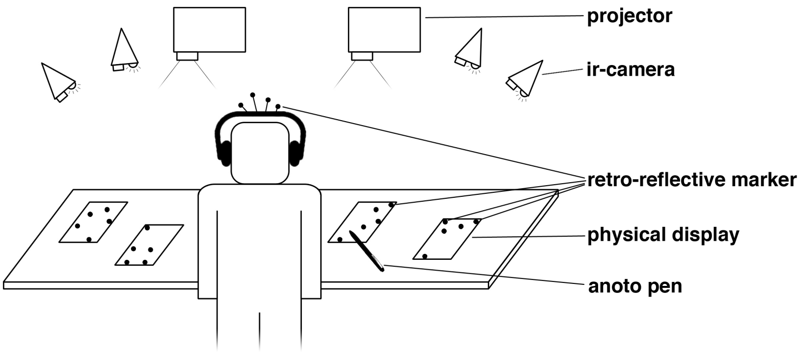


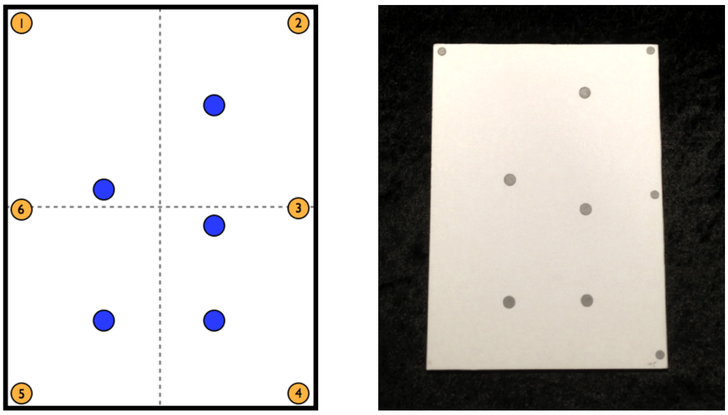
Figure 3.15: Technical setup.

We added a layer of Anoto pattern on each cardboard. By using Letras Framework, the pen could send pen coordinates, which we then converted to mouse events in the WPF application. We refrained from supporting touch input for the following reasons. Pressure-sensitive or capacitive touch foils either require tethering or too bulky of electronic components. While optical touch tracking would be a suitable approach with a single display or a small number of displays, it is too unreliable with the large number of displays supported by our system and the corresponding large number of markers, which is required for tracking of displays.

3.6.1.3 Sound Output

For generating, a 3-D perception of sound, we used the OpenAL Framework [OpenAL, 2010]. The user was equipped with headphones that where augmented with markers to track the user's head position and orientation so that the sound sources could be positioned accurately in space. With this implementation, we have implemented sound concepts to mentally grasp multiple audio sources of videos simultaneously. This concepts are in detail presented in chapter 4.

The first implementation was designed for single-user setting only. This implementation was sufficient for a single-user evaluation presented in subsection 3.7.1. In order to evaluate the system in a collaborative setting, the first implementation has been enhanced. The corresponding changes are presented in the next section.



(a) Blue markers in the middle of the display are used to track displays position and orientation. Orange markers afford display identification.

(b) A real display with the ID 15. Because orange markers that are availability are summed up using the binary system. $2^0+2^1+2^2+2^3=15$

Figure 3.16: Specific marker template is used that can be alternated to identify multiple physical paper-like displays.

3.6.2 Multi-User (Collaborative) Extension

CoPaperVideo is based on the same simulation environment as for the first prototype. We changed the following components of the framework in order to use our system collaboratively: (1) tracking algorithm of multiple displays, (2) rendering of simultaneous playing videos, (3) touch input, and (4) 3-D sound for multiple users. In the following, we will present the changes step by step.

3.6.2.1 Tracking of Multiple Displays

We have developed an algorithm that is based on predefined position of multiple passive infrared markers onto the cardboard. A group of markers is used to identify the display and its orientation and position in 3-D space (Blue markers in Figure 3.16a). Another group of markers allow to space unique identifier onto the physical display (Orange markers in Figure 3.16a). This algorithm allows us to distinguish more then 30 displays in parallel from each other in 3-D space.

A pseudo-code (see Algorithm 2) shows how we identified the displays based on the 3-D points provided by the simulation system.

Algorithm 2 Algorithm for tracking and identifying displays

```
procedure TrackDisplays(pointTree, displays))                          ▷ see (1)

    for each display in displays do
        points ← findDisplayPoints(pointTree, display)
        if points == null then
            remove display from displays
        else
            remove points from pointTree
        end if
    end for

    newDisplays ← detectNewDisplays(pointTree)                          ▷ see (2)
    for each display in newDisplays do
        add display to displays
    end for

    for each display in displays do
        neighbors ← calculateNeighbours(displays, display)             ▷ see (3)
        visibleID ← determineVisibleID(pointTree, display)             ▷ see (4)
        for id = 1 to MaxID do                                          ▷ see (5)
            rating ← rateID(display, id, neighbors, visibleID)         ▷ see (6)
            add (rating, id, display) to ratingSet
        end for
    end for

    sort(ratingSet)                                                     ▷ see (7)
    for each rating in ratingSet do
        if !(assignedIDs contains rating→id) and !(assignedDisplays contains rating→display) then
            assignID(rating→display, rating→id)
            add rating→id to assignedIDs
            add rating→display to assignedDisplays
        end if
    end for
    return displays
end procedure

(1) Displays serve as input (displays from last frame) and output (displays from current frame) set.
(2) Tries to find the trackable pattern with a backtracking algorithm.
(3) Calculates the neighbors, based on the spatial distance.
(4) Checks which points from the id pattern are assigned in the pointTree.
(5) MaxID is the maximum id that a display can have.
(6) Determines the probability that the given display has the given id, based on the neighbors and visible id.
(7) Sorts the set according to the rating.
```

3.6.2.2 Rendering of Simultaneous Playing Videos

The previous implementation of video rendering was designed in a restrictive way, having the following limitation: (1) only up to four videos could be played back simultaneously and (2) videos were rendered on two projectors only. In order to use CoPaperVideo with multiple users, both limitations have been improved and are described in the following.

In order to visualize and play back multiple videos simultaneously, videos have to be natively rendered on the graphic card. We explain the implementation by showing the class diagram (see Figure 3.17).

The *VideoSingleDevice* class creates a *FilterGraph* that allows to control the video, for example, play, pause, stop, and seek. It also creates an instance of the *VideoMixingRenderer9* class, which prepares the rendering of the video. Both classes are instances from the *DirectShow* library that allow an upload of the video in the memory of the graphics card. At the same time, while instantiating *VideoMixingRenderer9*, an instance of *VideoRenderer* is forwarded. *VideoRenderer* is responsible for the actual rendering onto the display. *|VideoRenderer* receives the needed frames of the video from *VideoMixingRenderer9* while the video is played. Information such as position, orientation, and corresponding display are combined with images of video frames and stored in the *TextureStorage*. Stored information is from there forwarded to the projector for rendering. In this way, the frames of the video go to the projector. Since rendering of videos happens in the graphic card, our implementation supports more then six simultaneous playing videos. By instantiating one *VideoSingleDevice* for each graphic card, multiple projector can be supported (see Figure 3.18). Hereby, all graphic card needs to store a texture representation of a single video to be able to render it.

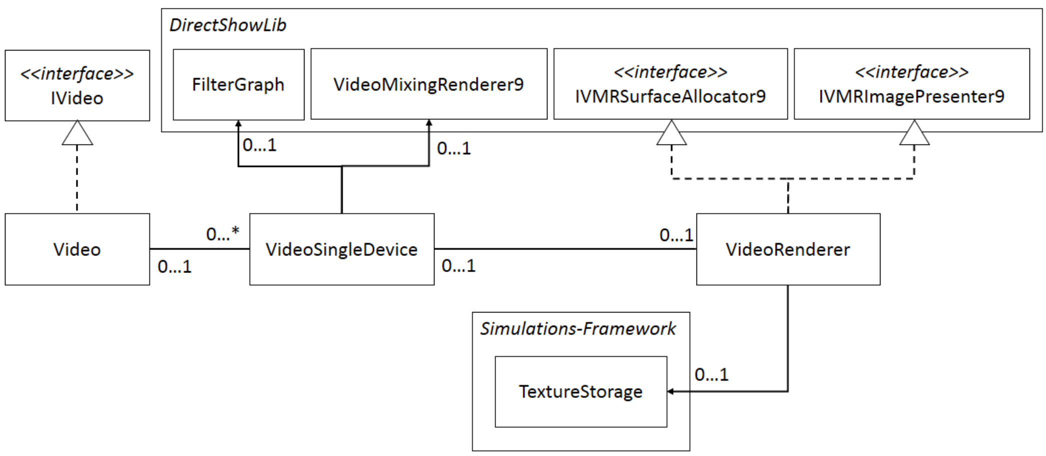


Figure 3.17: Video rendering - class diagram.

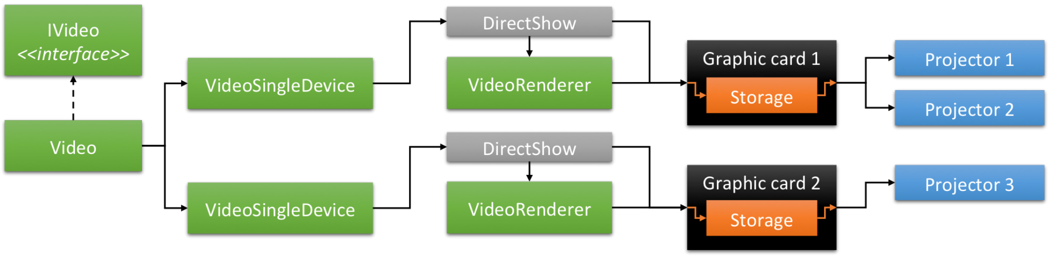
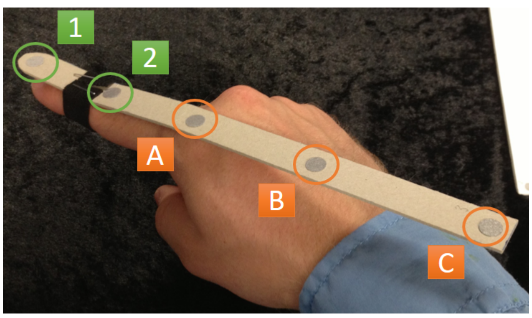


Figure 3.18: Rendering of videos on multiple projectors.

3.6.2.3 Personalized Touch Input

In order to allow a direct manipulation on the display, we have to augment the user with a cardboard stripe to allow touch input. IR markers are placed on the stripe marked as A, B, C in Figure 3.19a. These markers are combined to a line in 3-D space. Intersection of the line with the display with a specific angle and distance creates the touch point. Two additional markers 1 and 2, visible in the figure above, are used to distinguish between different stripes of different users to allow personalized touch input. Personalized input is needed for collaborative sound concept called *touch-based focusing* that is presented in section 4.5. Figure 3.19b shows the working touch prototype.



(a) A cardboard stripe on the finger allowed personlized touch.

(b) The working personlized touch.

Figure 3.19: Implementation of personalized touch.

3.6.2.4 Multi-User 3-D Sound

In order to evaluate our system collaboratively, each user needs to hear the sound of videos in 3-D space. Therefore, we tracked and augmented each user's head with stereo headphones as well as provided for each user a virtual 3-D sound output. A detail explanation of our implementation is described in chapter 4 in subsection 4.4.2.

3.7 Evaluation

We have evaluated our system in two iterative evaluations. First evaluation focused on a single-user scenario whereas the second evaluation focused on multi-user scenario. In the following both, evaluations are presented sequentially.

3.7.1 Single-User Evaluation

Overview – Study Design	
Method:	Qualitative evaluation
Interest in:	Users' practices, Mental models
Participants:	6 experts (single-user session)
Duration:	avg. 3h
Data gathering:	semistructured interview, observation and video-taped

CoPaperVideo introduces a novel way of tangible interaction while actively working with videos. It allows for a broad range of new styles of working with videos. Rather than focusing onto single variables (like time efficiency) and thereby limiting our view to a subset of scenarios, it is of primary importance to understand the broad range of new styles of working with videos that are enabled by our system. In particular, an evaluation must provide first insights into how users treat and use multiple displays simultaneously and how this affects the interaction with videos. Hence, answering the "why" and "how" is very important; this requires a qualitative research methodology, rather than a quantitative one, and a detailed analysis of a small but focused sample. The in-depth analysis of users' practices and mental models that are presented in this section provides the foundation for future quantitative analyses of specific questions.

3.7.1.1 Study Design

Six experts participated in 3-hour single-user sessions. All of them are video power users, spending large amounts of time on watching videos (avg=13 hours a week, SD=4.2) and having extensive recording and video authoring experience. To ensure a wide range of viewpoints, we recruited participants with different professional backgrounds: medicine, computer science, philosophy, cultural science and product design. All participants were male with a median age of 24. All of them were familiar with modern computing devices, all owning a smart phone and two of them owning an iPad.

Each session was organized as follows. After 10 minutes of guided introduction to the system and its basic interaction techniques, the participant was given ample time to get familiar with the devices. Then the participant was asked to perform the following tasks using a think-aloud protocol:

1. The first task consisted of navigating within one video on a single display. The participant was asked to give a short oral summary about two different scenes of the video: one situated near the beginning, the other near the end of the video. He was free to decide which interaction technique to use (pen or timeline in space)

2. The participant was given all seven displays. His task was to explore a collection of six related videos, to group them in two meaningful groups, to explain his grouping, and to regroup all videos following another criterion.

3. Before the third task, we introduced the interaction techniques for combining and distributing content. The task consisted of exploring a collection of 14 related videos and grouping it into two given topics, followed by giving an overview of all videos in the collection. This task was performed twice with different contents: once using all seven displays, and once using only three displays (the order was counterbalanced).

4. Before the next task, we introduced the lightweight video editing technique to the participant. He had to provide a summarized video of soccer goals by creating a video excerpt which contained the goal sequences from three soccer videos.

These tasks allowed us to study a range of phenomena: how users leverage space to interact with multiple videos on multiple displays (tasks 1-3), how they combine and distribute content over different amount of displays (task 3), and how they perform lightweight video editing (task 4).

An additional task that focused on how they were able to manipulate and focus on multiple parallel sound sources is presented in the next embodied sound interaction (see chapter 4 in subsection 4.4.3).

3.7.1.2 Data Gathering and Analysis

As methodologies, we used semistructured interviews (at the end of each task and after the whole session) and observation. The entire session was videotaped. Interviews and observations were transcribed and analyzed using an open coding approach [Strauss and Corbin, 2008].

3.7.1.3 Results and Discussion

Task Support

The system itself was positively received by all participants. For instance, P5 mentioned, *"I have the video with the page really in my hands (...) that's great."* All participants explicitly mentioned that they could easily gain an overview of the videos and could easily structure them by having multiple displays that could be rearranged in space. To more quickly get an overview of a video collection, one participant (P1) watched two videos in parallel on two adjacent displays. This allowed him to decide whether he liked the video and sort the video into the appropriate category.

Four participants emphasized that they would use the system for working purposes, such as learning, ordering, and organizing videos as well as video editing. However, these participants were skeptical in using the system to just watch videos, such as watching a theater movie or YouTube video. P4 stated that the system is *"good for teaching video navigation to PC novices."* P5 envisioned that the system be used at an exhibition booth in order to attract people via promotional videos.

We can summarize the results as follows:

- Easily gain an overview and structure videos
- Rearranging multiple displays in space creates a high sense of directness

Functional Zones

In the tasks where participants could use all seven displays, table space was used similarly to how it is reported in the literature about traditional paper documents [O'Hara and Sellen, 1997; Scott *et al.*, 2004]. In the video recordings, we clearly identified two different functional zones: The area situated directly in front of the participants, in the center of attention, can be characterized as the working area. Less important videos were moved to the periphery of attention, at the outer, more distant zones of the table surface. Following Scott *et al.*, we call this the storage area.

Most intense interaction with video displays (such as playback, seeking, accessing related videos, clearing contents, video cutting) was done in the working area. Only one single participant navigated videos in the storage space. In contrast, all participants moved displays that were currently not needed (whether they were filled with content or be they empty) out of their attention into the storage space. These displays were placed on a free spot or piled onto other displays. Three participants loosely arranged displays in the storage area, without piling. P6 explained: *"I have not piled because I'm not a person who orders things directly."* In contrast, other participants preferred less cluttered arrangements and piled the displays as soon as they placed them in the storage area.

In contrast to the above findings for seven displays, participants behaved differently in tasks where they disposed only of three displays. Two participants kept all displays inside the working area. Four participants placed only one single display in the storage area, which contained a virtualized pile. These findings show that the number of available displays directly influences the spatial practices of how the system is used. While with only three displays, the system is used very much like an enhanced computing device with only very limited spatial interaction, already a relatively small number of additional displays is sufficient for unleashing the power of paper-based interactions.

We can summarize the results as follows:

- We have identified two functional zones (storage area and working area) that have been previously reported in the literature about usage of traditional paper documents.
- With only three displays, the system is used very much like an enhanced computing device with only very limited spatial interaction. Most interactions happen in the working area.
- With seven displays users unleashing the power of paper-based interactions, hereby using both functional zones.

Functional Roles of Displays

As long as the number of videos did not exceed the number of available displays, all participants realized a fixed one-to-one mapping of videos to displays. The display has thereby one single functional role: being a physical carrier of the video content.

In cases where the number of videos exceeded the number of displays, we observed two general strategies of participants to cope with the fewer number of displays. One group (three of six) can be characterized as "**materializers.**" These participants preferred having as many videos as possible available in tangible form. They filled as many empty displays with content as possible. Only once all displays were full, they combined videos onto one carrier display to get free displays. In turn, all of these displays were filled before virtualizing again. In contrast, the other group (three of six) can be characterized as "**virtualizers.**" They filled just one (or at most two empty displays) and directly grouped the videos onto a virtual pile.

Participants assigned stable functional roles to physical displays. We identified *three different roles of display usage* that all participants assigned to displays in tasks 2–4: (a) information source, (b) working display, (c) information container. The information source was a display that contained all the related videos. The working display was used to iteratively open, watch, and assess a related video before moving it into an information container. An information container display was used to group and store several videos that represented a topic. Both "materializers" and "virtualizers" attributed these roles to displays, with the only difference being that virtualizers had only one working display, whereas materializers had several.

Despite the possibility of using many displays, P6 and P4, both "virtualizers," used only three or four of them. P6 stated, *"Oh yes that goes well with 3, with 3 I have a better overview of the displays."* P4 even felt three displays to be more efficient than with more of them and was amazed how easily the tasks could be performed. In contrast, P1 and P5, both "materializers," found it inconvenient to work with only three displays.

We conclude from the results that depending on their strategy, users prefer more or less working displays. The system should provide enough displays to leave the choice to the user.

We can summarize the results as follows:

- Dependent on the amount of displays in relation to number of videos (content) available for the task, participants had different working strategies.

- *# Displays > # Content*: Fixed one to one mapping, Display had one functional role being a physical carrier of the video content

- *# Displays < # Content*: Two different strategies were observed. (1) Materializers: Participants preferred having as many videos as possible available in tangible form. (2) Virtualizers: Participants filled just one (or at most two empty displays) and directly grouped the videos onto a virtual pile.

- We identified three different roles of display usage (a) information source, (b) working display and (c) information container.

Dynamic vs. Static Content

In this section, we focus on how dynamic content on paper displays was perceived in contrast to traditional static content on paper. Five participants found that dynamic content was easy and intuitive to use. P1 stated, *"It is better that videos can be detached from the medium whereas content on paper is bound to paper."* Furthermore the *"x-ray view"* (P5) through the pile was found to be beneficial to paper by all participants: *"I can see through the pile and still interact with it."* (P2). P5 mentioned, *"It saves me from flipping through pages. That's convenient; I can directly continue to work."*

All participants positively perceived that contents of the displays within a pile change dynamically when a pile is virtualized. P6 commented, *"This allows me to work with less displays."* However, comments about how specifically display contents should change in this case revealed two different mental models of the participants. This is best made clear by an example given by P5:The participant wanted to add a video to an existing virtual pile of videos. To do so, he placed the display with the single video on top of the virtual pile and then performed the up gesture for virtualizing the pile. This automatically "pushed" all contents to the topmost display, which now contained the virtual pile. The remaining displays were empty. Thereby, the former working display changed its role to an information container and vice versa. This participant had a mental model that focused on the roles of the physical displays. He disliked their changing roles, which he described as *"computer logic"* (P5). In contrast ,the remaining five participants had a mental model that focused rather on the dynamic contents. They did not even notice that the physical carrier medium changed its role.

From these results, we conclude that dynamic content on physical displays is appreciated and does not need high rethinking or high cognitive effort in contrast to the known behavior of traditional paper. However, to account for the mental model that focuses on the physical carrier medium, systems should equally support interactions that keep the roles of the physical displays steady. For instance, videos can be added to an existing pile by dragging and dropping them from the working display onto the information container.

We can summarize the results as follows:

- Dynamic content on physical displays is appreciated and does not need high rethinking or high cognitive effort in contrast to the known behavior of traditional paper.

- Systems should equally support interactions that keep the roles of the physical and digital displays steady. For example, moving digital content from one display to another should not force the user to move the display physically.

Physical vs. Touch Input

We were interested to find out which interactions should be delegated to physical input (manipulating displays in space) and which ones to surface-based input (direct touch or pen input).

We observed that interactions like arranging and piling videos in space as well as combining and distributing contents were performed intuitively by physical input. In contrast, all participants intuitively used surface-based input for playback and skimming in videos. The interaction for temporal navigation in space was rarely used. Participants stated that skimming with the pen requires less effort and less space and moreover is more precise than moving the display in space. Furthermore, two participants (P2, P4) proposed a function for reordering videos within a pile by dragging and dropping their small representations on the topmost display. P3 suggested that the list of related videos be accessed by touch and that related videos be opened by dragging and dropping them from the list onto an empty display. However, P2 appreciated the physical gesture and mentioned browsing relations, by bringing displays near to each other, is intuitive and easy to use.

From these results, we conclude that interactions that naturally anchor information in physical space would rather be done using physical interactions. For instance, if users arrange displays or pile them, they expect the physical arrangements to change. On the other hand, interactions that have no spatial anchor and are without spatial consequences are performed using surface-based input. This particularly concerns interactions that apply to only one single display, such as temporal navigation within a video. Future work should explore spatial temporal navigation with multiple displays that are linked to only one video. By moving and arranging displays in space, the user could sneak-peak into different temporal locations of the video simultaneously and easily compare contents within one video. We assume that in this case, users prefer physical over touch input.

We can summarize the results as follows:

- Interactions that naturally anchor information in physical space would rather be done using physical interactions.
- Interactions that have no spatial anchor and are without spatial consequences are performed using surface-based input.

3.7.1.4 Summary

We conclude the single-user evaluation by summarizing the evaluation results in Table 3.3.

3.7.2 Multi-User Evaluation

In this section, we describe our second evaluation that focuses on collaborative use of CoPaperVideo. In particular, in this evaluation ,we were interested in getting first insights into how *multiple* users treat and use multiple displays simultaneously and how this affects the interaction with videos. In the following, we present the study design, analysis methods, and study results. For this evaluation, with multiple users, we used our enhanced implementation of the system supporting multiple users. For implementation details, please read subsection 3.6.2.

Overview – Study Design	
Method:	Qualitative study
Interest in:	Multi-user practices
Participants:	9 experts (3 groups of 3)
Duration:	avg. 2h
Data gathering:	Semistructured interview, observation and video-taped

Summarized Evaluation Results:
– Our system allows users to flexibly organize and structure multiple videos in physical space.
– We have characterized different mental models and strategies of users (“materializers” vs. “virtualizers”) to cope with a restricted number of displays.
– We have identified different roles of display usage: (a) information source, (b) working display, and (c) information container.
– Depending on participants strategy, users prefer more or less working displays. The system should provide enough displays to leave the choice to the user.
– Dynamic content on physical displays is appreciated and does not need high rethinking or high cognitive effort.
– Interactions that naturally anchor information in physical space would rather be done using physical interactions. (e.g., arrange displays or pile them)
– Interactions that have no spatial anchor and are without spatial consequences are performed using display centric interactions. This particularly concerns interaction on a single display (e.g., temporal navigation within a video)

Table 3.3: Summary of results from the single-user evaluation.

3.7.2.1 Study Design

This study was conducted with three groups. Each group consisted of three users with an average age of 25 years. Five males and four females participated in the study. All users were familiar with the medium of video with an average time of watching videos per week of 14 hours. We chose a within-subject design with three tasks. In order to minimize learning effects, subtasks in task 1 were randomized. Tasks 1, 2, and 3 were designed in a way that they focused on different parts of the system so that learning effects were minimal. Participants did not have any time restrictions and decided on their own when the task was finished. At the end of each task, we interviewed the participants in a semistructured interview. Participants were standing throughout the whole study around a normal rectangular table. Paper-like displays could be placed on the table. The average study duration was 2 hours 20 minutes with an in-between break of 30 minutes.

Task 1: Sound

We had a single task that gave us an insight in how well previously presented single-user sound concepts will perform in a collaborative environment when multiple users try to work with videos on multiple paper-like displays. In order to address this question, we asked a group of users: *"Imagine you are planning a party and you are responsible for the music at the party. Please select out of 10 music videos your playlist for the evening. The playlist should have at least 3 music videos."*. This task was performed individually with all four different sound concepts. Results for this tasks are presented in the next chapter 4 in subsection 4.5.2.

After fulfilling the subtasks with all the four sound concepts, the participants had to decide their favorite concept. This sound concept was then used in the other task of the study.

Task 2: Search and Sort Videos

The following task examined multiple users collaboratively searching and sorting videos using CoPaperVideo. Our focus of observation was (1) how participants switched between individual and group work, (2) how they searched in a group, and (3) how they combined their results. The concrete task that we asked the participants was, *"You want to create a video collage of the following three actors: Brad Pitt, George Cloony and Julia Roberts. Please search among all presented movie trailers for movies they play in. Sort the movies you find by actor. Hint: Each starting video has related videos."*

Task 3: Video Editing

Commercial video editing is mostly done by only one person at a time. With this task, we wanted to see how multiple users would proceed while having the task to video edit one video out of multiple videos. We were particularly interested in whether participants will edit videos in parallel. The concrete task that we asked the participants was, *"All of you are working in the marketing agency of a film company. You were asked to generate a 30 seconds trailer for an upcoming release of a movie."*

3.7.2.2 Data Gathering and Analysis

For data gathering, we used semistructured interviews (at the end of each task and after the whole session) and observation. The entire sessions were videotaped. Interviews and observations were transcribed and analyzed using an open coding approach [Strauss and Corbin, 2008].

3.7.2.3 Lessons Learned

Collaborative Searching and Sorting of Videos

In this task, all three groups flexibly arranged and physically structured and play backed videos in parallel.

Participants divided their workspace in a private and group territory, as was previously shown on traditional tables with paper documents [Scott *et al.*, 2004]. Nearly all participants searched for videos individually in their private territory. Hereby users often used displays that were in the arm range of them. Final search results were summarized in the middle of the working area (known as group territory). One participant mentioned at the end of a task: *"Everyone has fought for himself and in the middle all the information was collected"* (P3). A similar working style was also observed in the other two groups.

Videos in the group territory were structured in two different ways: (a) combined on one display in form of a video collection or (b) piled on multiple displays. This structuring behavior can be reasoned with our findings in the first study, having different working strategies such as "materializers"[3] and "virtualizers."[4]

[3] Participants preferred having as many videos as possible available in tangible form.

[4] Participants filled just one (or at most two empty displays) and directly grouped the videos onto a virtual pile.

In order to synchronize their own working state, group members flexibly exchanged information in three different ways: (1) by showing the display to the other user: *"Is that Julia Roberts"* (P3) , *"Maybe this video is already in there?"* - *"No."* (P8 and P9); (2) by passing their display to the other participant, and (3) participants moved physically closer to each other to see the same content on the displays.

From these results we can conclude:

- Each user could flexibly arrange and physically structure videos.
- Participants arrange their working space in a similar way as they do with paper documents on a table.
- Participants followed the same working strategies such as "virtualizers" and "materializers" as known from the single-user study
- Participants flexibly exchange information and synchronize their working state in a similar way as they do with paper documents.

Collaborative Video Editing

During video editing task, we observed two different working strategies. All groups, however, assigned two different roles to participants such as *video cutter* or *video finder*. The video cutter role took control of cutting the videos. The video finder role focused mainly on searching for appropriate videos that could be passed to the video cutter.

Most groups worked with only one corresponding participant for video cutting. Other participants worked as video finders. The video finders worked in parallel, whereas the video cutter worked in sequential order. In only one group did the video cutter also search for video sequences in his idle time.

Some participants stated that they could imagine that our system could be used for small and private video editing projects because of its intuitive and easy way of physical interaction. One participant was skeptical whether this way of video editing could be used for professional use: *"for fun areas but for the professional area rather not."* (P1). Our proposed interaction techniques are in face only suitable for a rough pre-editing. A more precise frame-based video editing could be permitted by a combination of device-centric interaction and physical video editing.

From these results, we can conclude the following:

- We observed two different roles that were assigned to users during collaborative video editing such as "video cutter" or "video finder."

- Participants stated that our video editing interactions could be beneficially used in small or private video editing projects because of its intuitive and easy way of physical interaction.
- For a more professional and precise frame-based video editing, additional gesture, are needed.

Collaborative Spatial Sound

Most compelling results of this study were regarding collaborative use of spatial sound concepts. These results are ,however, presented in the chapter 4.

3.7.2.4 Summary

Table 3.4 summarizes our multi-user evaluation results.

3.8 Conclusion

In this chapter, we proposed a novel paradigm for users to spatially interact with video content. We introduced CoPaperVideo, a collaborative environment for spatial interaction with videos on paper-like displays. CoPaperVideo is the first system that brings a fully functional video directly to paper. Based on the design space for multiple spatially aware paper-like displays, we introduced a set of interaction techniques that support

Summarized Evaluation Results:
– Our system allows multiple users to flexibly organize and structure multiple videos in physical space.
– Participants arrange their working space in a similar way as they do with paper documents on a table.
– Participants flexibly exchanged information and synchronized their working state in a similar way as they do with paper documents.
– We observed two different roles that were assigned to users during collaborative video editing such as "video cutter" or "video finder."
– Participants stated that our video editing interactions could be beneficially used in small or private video editing projects because of its intuitive and easy way of physical interaction.

Table 3.4: Summary of results from the multi user evaluation.

playback, flexible navigation and spatial organization of videos on multiple physical displays.

Results from two iterative evaluations gave first insights into how users treat and use multiple displays simultaneously and how this affects the interaction with videos. (1) Results from *single-user evaluation* show that users can flexibly organize and structure videos in physical space while generating a good overview of multiple videos. They thereby flexibly attribute three different functional roles to paper-like displays: information source, working display, and information container. We have also characterized different mental models and strategies of users ("materializers" vs. "virtualizers") to cope with a restricted number of displays. (2) *Multi-user evaluation* results indicate that CoPaperVideo is suitable for collaborative use. CoPaperVideo allows users to work individually as well as in a group while allowing to transition between coupling styles in a similar way as with paper documents. Furthermore, participants could flexibly exchange information and synchronize their working state in a similar way as they do with paper documents.

We have developed a simulation environment to support spatial interaction with multiple spatially aware displays. Even though our simulation environment allows multiple users to interact with multiple paper-like displays, our evaluation results for the collaborative scenario is limited to three groups of three users. Further quantitative evaluations are needed to draw general conclusions about a user's mental model and interaction patterns.

In this chapter, we have presented the following:

- CoPaperVideo, a collaborative environment for interaction with videos on multiple spatially aware paper-like displays.
- We contributed a design space for spatially aware paper-like displays.
- We introduced a set of interaction techniques that support playback, flexible navigation, and spatial organization of videos on multiple physical displays.
- We evaluated CoPaperVideo with two user studies focusing on single and multi-users interaction with the system.
- Results indicate that CoPaperVideo allows users to flexibly organize and structure multiple videos in physical space.
- In addition, CoPaperVideo allows users to work individually and in a group while allowing for flexible transitioning between coupling styles.

Videos consist out of two output channels: (1) visual output and (2) sound output. This chapter presented interaction techniques and visualizations to manipulate and control the visual output. The following chapter will focus on interaction with sound output for single and multiple users.

I would like to conclude this chapter by quoting Fitzmaurice [1993] and elaborating on the point that in this chapter we presented lightweight, spatially-aware paper-like displays for active video work. I believe, in 1993, Fitzmaurice had envisioned the displays we are currently still simulating in this chapter.

> “ [...] These displays are aware of their surroundings and change depending on the situation in which they are immersed. [...]
>
> *George W. Fitzmaurice in 1993* ”

4

Embodied Sound Interaction

Contents

In chapter 2, we previously presented how users collaboratively work on a novel interactive tabletop called Permulin and proposed interaction techniques that allowed for flexibly transitioning between individual and group work by providing full-screen private views. Hereby, we studied interaction effects and collaboration on a *single display*. In chapter 3, we introduced a novel concept for video navigation by introducing spatial interaction techniques for *multiple spatially aware paper-like displays*. However, until now, our interaction and device concepts only focused on **vision-based** output modality. In this chapter, we will present how to support fluid collaboration with another modality such as **sound-based** output, hereby emphasizing hearing as another important human sense for interaction.

In order to emphasize why interaction with sound is so inspiring and crucial for us, we would like to mention this quote that compares the visual and auditive human sense:

> “ The ear differs from the eye in that it is omni-directional – a true three-dimensional (3-D) display space that does not suffer from occlusion. Its fabric is coarse grained – with angular resolutions approximately 10 times more coarse than the eye across the sensorally richest regions [Howard and Templeton, 1966].
>
> *Walker* et al. *[2001]* ”

Although sound perception has the above stated advantages, direct manipulation [Hutchins *et al.*, 1985] of digital contents has become indispensable after introducing touch-based displays because of its intuitive and natural mapping that is inspired from the real world. Direct manipulation is also suitable for collaboration in a group of people specially during co-located collaboration as shown in chapter 2.

Direct manipulation of multiple sound sources is, however, still an unsolved challenge, specially when focusing on individual and group work with sound. Commercially available products provide, for example, direct playback of sound on the mobile device or indirect control of sound on speakers with a mobile device. Both, however, lead to a collaborative sound output heard by all people around these devices. Individual sound can be played back with earphones, allowing for listening to individual sound. However, also here, the interaction with this sound can only be done indirectly through a remote control on the earphone cable or through the mobile phone. Devices that allow users to manipulate individual or group sound as well as easy and intuitive switch between individual and group sound in a direct manipulative way, as we know it from visual-based interfaces, are to our knowledge not existent.

In this chapter we contribute two novel ways of interacting with sound in a direct manipulative way. First, we contribute a direct way of interacting with sound individually. Second, we present spatial interacting with sound that allows a collaborative way of directly interaction with sound individually or in a group of people. We advocate that both interaction styles are needed to control sound directly and permit users for fluid collaboration with sound.

Individual sound is perceived through our ears. Based on that, we propose to touch the own ear for **directly interacting with sounds individually**. Thereby, we contribute a novel device concept, called EarPut. It augments users earphones, and unobtrusively instruments the ear as an interactive surface. This allows each user to directly control their individual sound by touching their own ear. This body-based interaction has additional advantages. The human sense of proprioception enables us to relatively position our own body parts to each other without looking at them. Thus, a user does not necessarily need a visual interface for interaction and can interact eyes-free. Furthermore, the human ear is easy accessible for single-handed or bimanual interactions.

Sound is perceived omnidirectional, whereas the sound source is mostly placed in the 3-D space. In order to allow users a **direct manipulation of sound in an individual and in a collaborative way simultaneously**, we place sound virtually in 3-D space. This virtual sound is physically associated with spatially aware paper-like displays that embed videos. By physically moving these displays, each user can then control and focus on multiple sound volumes. This happens in a similar way as we are used when structuring and organizing physical objects in the physical world (e.g., documents or books). The proposed spatial interaction contents provide each user the ability to focus on multiple sound sources individually and collaboratively, while being able to fluently switch between them.

In summary, in this chapter ,we focus on novel devices and interaction techniques that allow sound output and control of sound in an individual and collaborative way as well as fluently transitioning between both collaboration styles; more precisely, we contribute the following:

1. We present how the human ear can be used for ear-based input to *privately* control sound. Allowing *individual* manipulation of sound, we present a novel device, called *EarPut*. Augmenting accessories that are placed or worn behind the ear such as behind-the-ear earphones or headsets with the EarPut device that unobtrusively instruments the ear as an interactive surface. This allow the user to remotely control and manipulate sound source privately by touching their own ear.

2. We contribute several novel sound concepts that allow a *single* user to mentally grasp multiple audio sources simultaneously. Hereby we virtually place the sound sources in 3-D space and allow each user to control the volume by physically moving paper-like displays in space.

3. An iterative design process with two user studies provided spatial interaction techniques with multiple paper-like displays for *individually and collaboratively* controlling volume of multiple sound sources.

These two novel devices and interaction concepts combined, advocate a novel collaboration paradigm for sound, allowing simultaneous control of multiple sound sources in a group by direct manipulation as we used while interacting with visual content. In this context, we study how interaction and collaboration from both non-visual (eyes-free interaction in form of body-based interaction) and visual (spatial interaction) interaction can be transferred to the sound domain. Thereby, we presented contributions propose to control sound individually with either *body-based interaction* by touching the user's own ear or with *spatial interaction* by spatially moving paper-like displays in 3-D space. Both interaction styles combined are named **embodied** [Dourish, 2004], due to their body-centric and physical interaction style.

This chapter has partially been published at ACM SIGCHI Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI) [Lissermann *et al.*, 2014a], ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) [Lissermann *et al.*, 2013a], and ACM Multimedia [Lissermann *et al.*, 2012b].

The remainder of this chapter is structured as follows: section 4.1 first reviews related works in the area of body-based interaction and spatial sound that are highly related to embodied sound interaction. In section 4.3, we present EarPut, a device that augments accessories such as behind-the-ear earphones or headsets to allow sound control by touching the user's own ear. Furthermore, we present a controlled experiment that shows in how many areas a human can mentally divide their own ear and how many of these certain areas can be precisely and effectively touched by the user. In the next two sections, we introduce spatial interaction techniques and their evaluation for parallel sound control for a single (in section 4.4) and multi-user (in section 4.5) setting. Lastly, we conclude this chapter.

4.1 Related Work

In the following, we first present related work focusing on body-based interaction. Then, we present research projects related to virtually or physically spatially located sound. Then, we explain the technical realization of how sound can be positioned in 3-D space.

4.1.1 Body-Based Interaction

In this section, we present the following research areas: *around* and *on-body* interactions as well as ear-based interaction. In the following, we illustrate how body-based interaction can control sound and situate our contributions within this space.

4.1.1.1 Around-the-Body Interaction

Miniaturization is an ongoing trend to make mobile interaction a reality. One may expect that in the future, devices that are rather large today, will be sufficiently small to be worn by tomorrow's users. This trend is particularly relevant for wearable computing, such that devices are getting sufficiently small to be worn on the body. Various researchers mounted sensors on the body and open up opportunities for *around-the-body interaction*.

Various projects mounted cameras on shoes [Bailly *et al.*, 2012] or onto the chest [Gustafson *et al.*, 2010; Mistry and Maes, 2009] to be able to track body movements. In all cases, the hands of the users are tracked, which leverages hand-based interactions as input capabilities. OmniTouch [Harrison *et al.*, 2011] uses a depth-sensing camera and a pico projector that is placed on the user's shoulder to allow interaction on arbitrary surfaces. Another example is Armura [Harrison *et al.*, 2012a], a ubiquitous projection system that tracks a user's hands to provide input capabilities. Furthermore, it uses the hands as a projection surface. Others augmented devices users already wear, for example, rings [Ashbrook *et al.*, 2011], cords [Schwarz *et al.*, 2010] (e.g., for earphones), and clothes [Karrer *et al.*, 2011], with input capabilities.

4.1.1.2 On-body Interaction

Various research has focused on *on-body interaction*. Here, a user's body is instrumented as an interactive surface. Users then interact, for example, with projected content on the forearm [Harrison *et al.*, 2010]. Another example is a research thrust that focuses on so-called imaginary interfaces. Pioneering research has been undertaken by Gustafson *et al.*, who investigated how to map a phone UI to a user's palm [Gustafson *et al.*, 2011, 2013], as well as by Dezfuli *et al.*, who investigated palm-based imaginary interfaces for TV remote interaction [Dezfuli *et al.*, 2012].

uTrack [Chen *et al.*, 2013] allows tracking of 3-D position of a user's thumb with two magnets on the back of the hand and a magnet on the thumb. Amento *et al.* [2002] has introduced a wristband enhanced with a microphone that can sense the sound produced by the hand such as tapping, rubbing, and flicking. Wagner *et al.* [2013] has developed a body-centric design space and investigated the effectiveness of on-body input while pointing towards interactive walls. A profound literature overview can be found in Harrison *et al.* [2012a] and Gustafson *et al.* [2013]. Another research thrust has focused on enabling on-body interaction through implanting technology into the body for implanted user interfaces [Holz *et al.*, 2012].

The systems described above mostly require heavy instrumentation (e.g., implantation) or additional devices to leverage interaction capabilities. Our vision is to design and implement unobtrusive add-ons to existing devices that instrument the ear as an interactive surface with as little setup as possible.

A recent study by Weigel *et al.* [2014] has explored various skin-specific input modalities such as touch, grab, pull, shear, and others. Their elicitation study show how users perform different gestures or commands via skin input. The study of Weigel *et al.*, however, only focuses on skin input on the upper limb of the human body.

4.1.1.3 Ear-Based Interactions

There is only a relatively small amount of previous works that focused on *interaction around the human ear*. Earphones have been enhanced in order to recognize hover gestures [Metzger *et al.*, 2004] or touch input on the headphones [Buil *et al.*, 2005]. Blindsight [Li *et al.*, 2008] investigated back-of-device interaction with mobile phones allowing eyes-free interaction around the human ear. One exemplary application is to allow users to access their calendar with the mobile phone buttons while holding the phone upside down during phone calls. Whisper [Fukumoto and Tonomura, 1999] is a wrist-worn handset compound of a microphone for voice input that also transmits a sound signal from the wristband to the users finger. The user can then listen to the sound by placing or even "plugging" the finger into one's ear.

None of the above projects leveraged the human ear as an interactive surface, for example, for touch-based interaction. However, the human ear exhibits unique affordances that we believe to be highly beneficial for a variety of applications:

Proprioception: The human sense of proprioception [Sherrington CS., 1906] enables us to relatively position our own body limbs to each other without looking at them. Thus, a user does not necessarily require a visual interface for on-body interaction. This particularly holds for the human ear.

Natural Tactile Feedback: The mechanoreceptors in the human skin provide means for immediate natural tactile feedback [Montagu, 1986]. This applies to both finger and ear during touch-based interaction on the ear.

Eyes-free Interaction: The prior observation leads to eyes-free interaction: Using eyes-free interaction has major advantages in the following categories [Yi *et al.*, 2012]: environmental (e.g., allowing interaction under bad lighting condition or improve safety in task-switching), social (e.g., avoiding interruption to social activities), device features (e.g., enable operation with no/small screens), and personal (e.g., lower perceived effort).

Easy Access: The human ear is easily accessible for single-handed or bimanual interaction. Single-handed interaction is particularly relevant in mobile settings, when it cannot be taken for granted that a user has both hands to her availability. The human ear is also one part of the body, which stays mostly uncovered by cloths.

These three observations motivated us to contribute a device that augments the human ear and allows humans to interact on their own ear arc to control sound output. Further details of our contribution are presented in the following section 4.3.

4.1.2 Spatial Sound

Multiple parallel audio outputs located in space produce the well-known cocktail party effect [Arons, 1992]. This spatial sounds can be fade away or maintained in focus by humans. This ability helped to develop interactive auditory systems by using virtual 3-D sound or spatially located sound. Both research areas will be discussed in the following. Lastly, we present research projects presenting technical realizations to place virtual 3-D sound.

4.1.2.1 Virtual 3-D Sound

In the following, research is presented that focuses on spatial sound. This virtually located sound is perceived by the user in 3-D space. From here on, we call this virtually spatially located sound: *3-D sound*.

In 1995, one of the first projects that tried to separate a mix of sounds into multiple spatially separated sound sources was Audiosteamer [Schmandt and Mullins, 1995]. Three years later, Audio Hallway [Schmandt, 1998] presented a spatial browser for related sound sources. An auditory interface could position sounds in a way that a user could spatially memorize and navigate and recall multiple sound sources without visual cues. A similar approach with a different application was presented by Walker *et al.* [2001]. He used an auditory interface that allows a user to access audio description of daily appointments via spatially placed sound sources around his head (see Figure 4.1). In 2007, Earpod [Zhao *et al.*, 2007] introduced a spatial virtual audio menu that could be browsed on circular touchpad. Sound feedback of the selected item was located at the same location as the user touch position on a circular touchpad. Their study indicates that Earpod is more efficient compared with standard visual selection with the touchpad after a 30 min usage time. Another work by Healy and Smeaton [2009] presented also a spatially aware audio playback that could illustrate auditively a virtual zoo for a user.

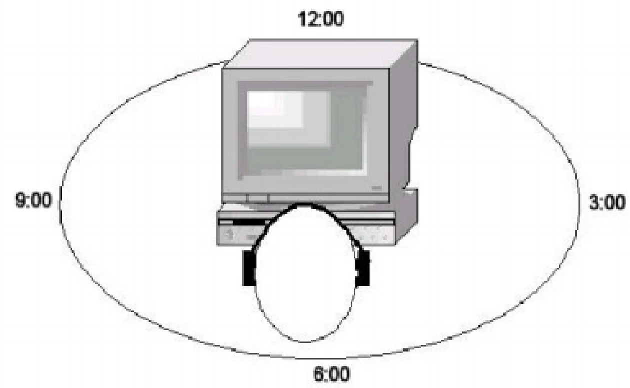


Figure 4.1: Daily auditive appointment descriptions are placed around user's head on a $360°$ virtually projected timeline. Figure was taken from Walker *et al.* [2001].

In 2010, Yolanda studied the efficiency and workload of a user when he wants to listen to simultaneous sound [Vazquez Alvarez and Brewster, 2010]. She compared interrupting and sequentially playing two sound sources with placing the sound virtually on different 3-D locations around the user. Her results show when focusing on simultaneous sound, moving both sound to a different virtual location in 3-D space can be promising. She also could show that doing so makes simultaneous audio presentation more usable. However, it can be less effective under high cognitive load. Furthermore, Vazquez-Alvarez *et al.* [2012] introduced an auditory display and studied how spatial sound can help users to explore an environment such as a park. Results of this study indicated that 3-D sound navigation was best suited to guide the user in an audio-augmented reality. Furthermore, she focused on understanding how many sound sources users can perceive eyes free. Her results show that at most two simultaneous sound sources can be heard virtually Vazquez-Alvarez and Brewster [2011].

These works presented interaction techniques for browsing multiple synchronous audio sources by creating an *audio-only* environment and virtually placing the audio sources around the user's head. By touching a device or simply turning the head, the user could select his sound of interest. In our work, we refine this research for audiovisual contents and contribute tangible interaction with sound while introducing further spatial audio concepts.

4.1.2.2 Spatial Interaction with Sound

Sound was placed in physical space and made accessible for multi-touch interaction by previous research projects. In the following, we present these works and elaborate on their contributions.

Morris studied how users collaborate around a shared tabletop while each user had a private audio channel [Morris *et al.*, 2004]. Their system is depicted in Figure 4.2. Each user could access their sounds by touching the surface. A Diamond Touch interactive tabletop was used that allows a personalized touch input for each user [Dietz and Leigh, 2001]. Morris study results revealed that it affects the collaboration strategies; however, it does not prevent group communication and can positively impact group dynamics. This work motivated us also to use earphones during our embodied interaction with sound.

Multi-Audible Table [Kusunoki *et al.*, 2004] introduced an interactive tabletop that allows each user to hear a digital content by touching it with the finger. The system hereby works as follows: an infrared light is transmitted from the table to the user's finger. The finger is augmented with a small solar cell that can encode the received light, so that the system knows who is touching the corresponding spot. Earphones that are worn by the user and connected to user's finger solar cell, play the corresponding audio source privately through the earphones. This systems builds upon previously presented work, which presented how information can be transported through light [Nakamura *et al.*, 2003]. Another interactive tabletop called Tangoscope [Edelmann *et al.*, 2011] supports parallel audio output privately for each user on a shared interactive surface (see Figure 4.3). A user can hear a specific digital content visualized on an interactive tabletop by positioning the stethoscope head onto the digital video/audio source. The stethoscope head is hereby augmented with earphones and is recognized by the tabletop. Results of a user study showed that while using Tangoscope users needed less time to start the first playback, they also played sound sources more often and continued listening to the sound longer. Both of the previously mentioned works, however, do not focus on support of multiple synchronously playing sound sources, as is the focus of our scenario.

Figure 4.2: An interactive tabletop that allows each user to have their private auditive information played trough their individual earbuds. Figure was taken from Morris *et al.* [2004].

Figure 4.3: Each user accesses individual sound sources on a shared tabletop using a stethoscope head. Figure was taken from Edelmann *et al.* [2011].

Jordà [2010] presented a music performance by using multiple tangibles on a tabletop that were controlling different synthesizers and sounds of instruments that were combined to a single music song. In this project, markers were tracked using reacTIVision fiducial markers [Kaltenbrunner, 2009; Kaltenbrunner and Bencina, 2007]. The song was played to all users and thereby did not provide an individual sound to each user.

4.1.2.3 Technical Realization of Spatial Sound

Different techniques to position sound in 3-D space exists. They can be distinguished by placing the sound virtually in 3-D or physically placing a sound source at a location in the 3-D space (e.g. a room). Both techniques are well known and are shortly described in the following.

Measuring and the use of head-related transfer functions (in short HRTFs) have been introduced by Gardner [1999]. He presented a technique for recording and positioning 3-D sound *virtually* with the use of headphones. Hereby, 3-D sound is recorded using HRTFs. The resulting database of records is used to calculate sound transformation characteristics for a particular reference point. In our case, the reference point is the user's head. HRTFs are then used to develop different finite impulse response (in short FIR) filters for each ear. These filters allow to place two monaural sounds, one for each ear though headphones, thereby allowing to position 3-D sound virtually around user's head (see Figure 4.4). A detailed description can be also found in Vazquez-alvarez [2013]. The same concept is used in our system. Today's audio frameworks provide support for 3-D virtual sound playback.

Another technique called Beam Forming [Van Veen and Buckley, 1988] allows directional positioning of *3-D sound in a specific 3-D location* (for example a room). These techniques,

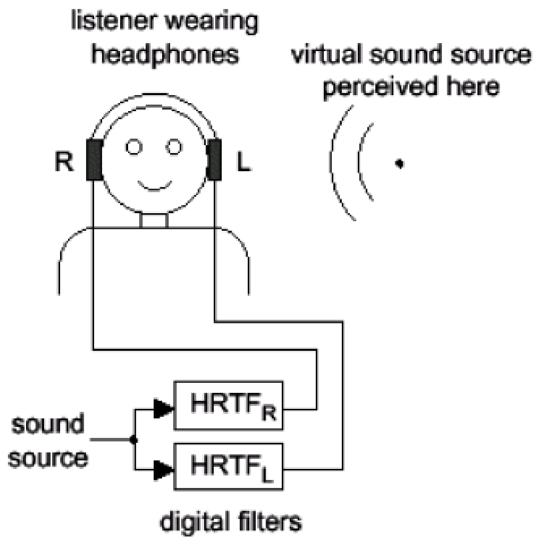


Figure 4.4: Placing virtual 3-D sound using HRTFs. Figure was taken from Gardner [1999].

however, are still not precise enough to use in a small space environment like a table as needed for our purpose.

We have presented related work in multiple research fields: body-based interaction, spatial sound, and technical realization of spatial sound. In the next section, we will present, how input on the human ear can enhance control and interaction with sound.

4.1.3 Guidelines and Summary

Based on the review of the previously presented related work, we present guidelines for designing sound interaction concepts that allow a single user and a group of users to control and manipulate sound in a direct way.

Interaction with Sound

GL1: Body-Based Sound Interaction In order to allow each user to control sound individually future system should support body based interaction. This allows each user to control sound privately on their own body.

GL2: Spatial Sound Interaction In order to control virtual 3-D sound collaboratively, future system should support spatial sound input. This allows each user to control the volume in 3-D space by for example physically moving paper-like displays in space.

GL3: Easy Accessibility of Controls Basic sound navigation controls should be easy and intuitive accessible.

GL4: Minimal Visual Attention Sound does not need users' visual attention. User interfaces should support this by allowing users to interact with individual sound with minimal visual attention.

GL5: Support of Virtual 3-D Sound Sound can be placed virtually into a 3-D space. This is crucial for allowing users to directly manipulate sound with spatial or physical interactions.

GL6: Support of Simultaneous Sound Playback An important requirement for getting an overview or comparing sound content is the ability to playback, listen, and manipulate multiple sound sources simultaneously.

Sound Output

GL7 - Individual Sound: If desired, each user should be able to personally focus on her (single or multiple) sound sources.

GL8 - Group Sound: If needed, a user or a group of users should be able to focus on a single or multiple common sound sources.

GL9 - Fluid Sound: Users should be able to easily switch between the individual and group sound mode.

In the following, we compare previously presented related work in respect to the guidelines (see Table 4.1).

4.2 Supporting Collaborative Concepts for Embodied Sound Interaction

Inspired from the direct manipulation on visual interfaces, our concept supports direct manipulation for auditory interfaces. Each user can manipulate and control multiple sound sources in a individual and collaborative way.

Individual sound is perceived through our ears. For direct interaction with the perceived sound, we propose a novel device concept, called EarPut. EarPut is a smart earpiece that augments accessories that are placed or worn behind the ear (e.g., earphones). It allows users to instrument their ear as an interactive surface to enable eyes-free mobile interaction with the ear. This allows each user to *directly control their individual sound by touching their own ear.* Since EarPut primarily focuses on input, we envision it as a

	Body-Based Sound Interaction	Spatial Sound Interaction	Easy Accessibility of Controls	Minimal Visual Attention	Support of Virtual 3-D Sound	Support of Simul. Sound Playback	Individual Sound	Group Sound	Fluid Sound
Guidelines	1	2	3	4	5	6	7	8	9
Speakers	○	○	○	○	○	○	○	●	○
Beam Forming [Van Veen and Buckley, 1988]	○	○	○	○	○	◐	◐	●	○
Conventional Earphones	○	○	◐	○	○	○	●	○	○
Audiosteamer [Schmandt and Mullins, 1995]	○	○	○	○	●	●	●	○	○
Audio Hallway [Schmandt, 1998]	○	○	○	●	●	●	●	○	○
Earpod [Zhao *et al.*, 2007]	○	○	◐	◐	●	○	●	○	○
Virtual Zoo [Healy and Smeaton, 2009]	○	○	○	○	●	●	●	○	○
3-D Sound Navigation [Vazquez-Alvarez *et al.*, 2012]	○	○	○	○	●	●	●	○	○
Private Audio Channels around Shared Display [Morris *et al.*, 2004]	○	○	○	○	○	●	●	◐	◐
Multi-Audible Table [Kusunoki *et al.*, 2004]	○	○	○	○	○	○	●	◐	○
Tangoscope [Edelmann *et al.*, 2011]	○	◐	○	○	○	○	●	◐	◐
The Reactable [Jordà, 2010]	○	●	●	●	○	○	○	◐	○
EarPut	●	○	●	●	○	○	●	○	○
Spatial Interaction Concepts	○	●	●	◐	●	●	●	●	●
Collaborative Concept for Embodied Sound Interaction	●	●	●	●	●	●	●	●	●

Table 4.1: Comparison of related work with collaborative concept for embodied sound interaction ● : completely fulfilled requirement. ◐ : partially fulfilled. ○ : not fulfilled.

The guidelines are derived as follows: Interaction with Sound: GL1, GL2, GL3, GL4, GL5, and GL6. Sound Output: GL7, GL8, and GL9.

companion device that piggybacks onto existing feedback mechanisms, for example, to wirelessly trigger auditory or vibrotactile feedback through actuators of a smart phone.

In addition to augmenting the ear for ear-based sound interaction, we propose spatial interaction with multiple sound sources for individual or collaborative direct manipulation of sound. Hereby sound is virtually placed in 3-D space and allows each user to control the volume by physically moving paper-like displays in space. The proposed spatial interaction with sound allows each user to control sound in a direct way as we do while structuring and organizing paper documents. Our spatial interaction provides each user the ability to focus on multiple sound sources individually and in a group of people.

The presented contributions are controlling sound either through *body-based sound interaction* by touching the user's own ear or via *spatial sound interaction* by spatially moving paper-like displays in space. Both interaction styles are **embodied** and are combined to support a fluent switch between individually and collaboratively interacting with sound. Hereby both novel sound interaction concepts can function in combination with each other (see Figure 4.5). Users are wearing earphones in order to have virtual 3-D sound that is needed for allowing spatial interaction. These earphones can unobstrusively be augmented with EarPut to allow each user a private direct sound control.

In the following, we compare our presented concept with our previously introduced guidelines (see Table 4.1).

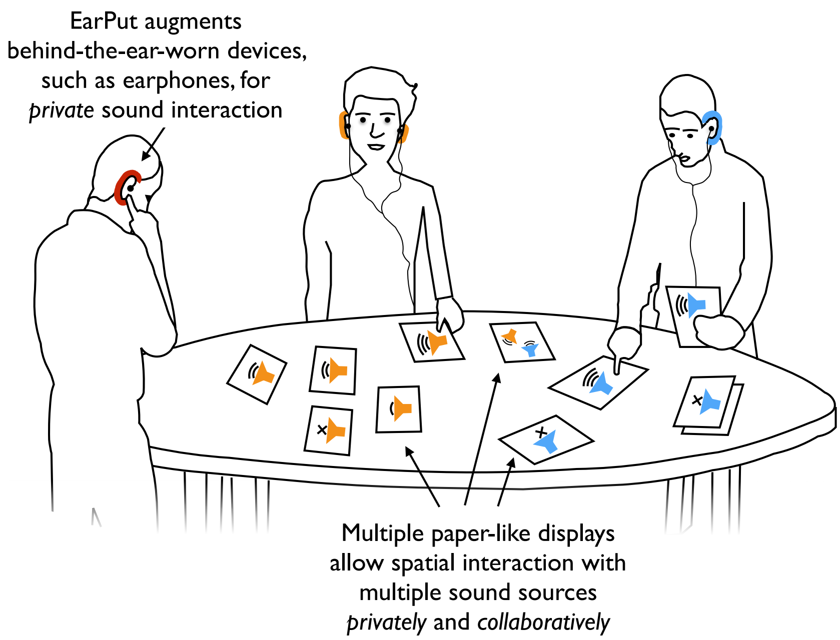


Figure 4.5: The concept of embodied sound interaction allows users to have a direct interaction with sound individually or in a group.

In the next section, we first present sound concepts for single-user interaction by directly touching the user's own ear to control sound. Secondly, we present single-user interaction techniques for directly controlling virtual 3-D sound by spatially moving paper-like displays (see section 4.4). Hereby each paper-like display can feature a video including the sound. Lastly, we present an iterative design process for creating collaborative sound concepts that support fluid collaboration with sound (see section 4.5).

4.3 Single-User Ear-Based Sound Control

Individual sound is perceived through our ears. For direct interaction with the perceived sound we propose a novel device concept, called EarPut. EarPut is a smart earpiece that augments accessories that are placed or worn behind the ear for example earphones. It allow users to instrument their ear as an interactive surface to enable eyes-free, mobile interaction with the ear. This allows each user to *directly control their individual sound by touching their own ear.*

Ears are particularly interesting for eyes-free mobile interaction due to three main reasons:

(1) We can interact with each of our ears using just one hand.

(2) The human sense of proprioception [Sherrington CS., 1906] allows us to do so reliably without visual attention.

(3) The ear as an interactive surface provides more degrees of freedom for interaction than, for example, ear- or headphones with integrated controls.

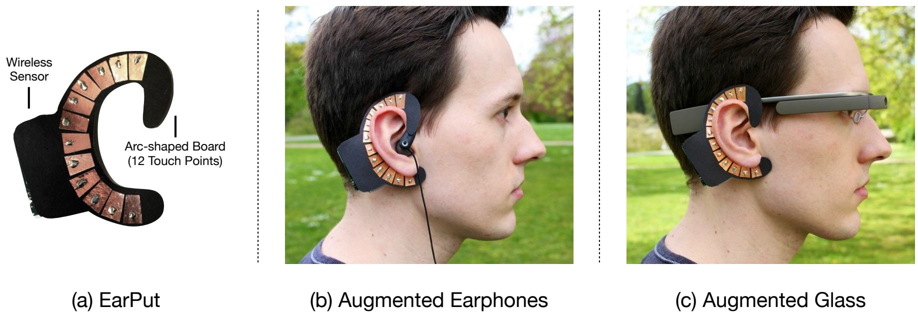


Figure 4.6: (a) EarPut augments ear-worn devices and accessories. It instruments the human ear as an interactive surface. (b) It serves as an interaction enabler for otherwise non-interactive devices such as ordinary earphones. (c) Also, EarPut complements existing interaction capabilities of head-worn devices, serving as a touch-based extension to e.g. Glass' touch-enabled frame.

These observations lead to the central question: *How can the specific characteristics of the human ear be capitalized for precise and effective eyes-free, mobile interaction?* With this question in mind, we propose to augment accessories that are placed or worn behind the ear such as glasses, ear hook earphones or headsets (see (see Figure 4.6)) with a device that unobtrusively instruments the ear as an interactive surface. We call this novel interface concept *EarPut*. This way, arbitrary ear-worn accessories can be augmented to enable eyes-free, mobile, touch-based interaction on the ear.

The following results of a controlled experiment are presented to assess both precision and effectiveness of touch-based interactions on the ear. The experiment lays ground towards more complex interactions. We then systematically derive a first design space for ear-based interactions that comprises interaction primitives and interaction syntax. Finally, we discuss the design and hardware implementation of EarPut and showcase implemented applications.

4.3.1 Preliminary Study

Overview – Study Design	
Method:	Quantitative evaluation
Participants:	27 participants (single-user session)
Duration:	avg. 15 min
Data gathering:	Precision and effectiveness of single touch and semistructured interviews

Independent Variables
Amount of areas, considering region-based interfaces with 2 to 6 different equally sized areas

Dependent Variables
Success rate of a user touching the highlighted region on her ear arc

The sense of proprioception allows us to reliably touch our own ear. However, it is unclear how (1) precisely and effectively users can touch certain areas, and equally important, (2) how many different areas can be targeted at all. We investigated these questions in a controlled experiment with 27 participants. The apparatus used in the experiment allowed us to measure both precision and effectiveness of single-touch interactions on the ear as a crucial basis for more advanced interactions. Moreover, we conducted semistructured interviews to obtain qualitative user feedback.

4.3.1.1 Apparatus

In order to track and identify touch-based interactions with the ear, we used capacitive sensing based on electrodes that are placed onto an arc-shaped area. Touch recognition was based on the MPR121 Capacitive Touch Sensor [Semiconductor, 2010]. We used the same chip also used in our second iteration of the hardware design (described later in this chapter). All 12 electrodes of the sensors are connected to 12 distinct areas on the arc touch areas. When either a finger or parts of the ear approaches an electrode, the electronic capacity increases, which is detected by the MPR121. Since we did not design a PCB in our first iteration, we used a breakout board of the MPR121 sensor. The combined device (i.e., the electrode arc and the touch sensor) was then used to augment the ear hook of existing wearable accessories (see Figure 4.7), allowing for touch-based interactions on the ear arc (i.e., on both ear helix and lobe). The breakout board is connected to a Arduino system, which reads the sensor data and forwards them to the computer via USB.

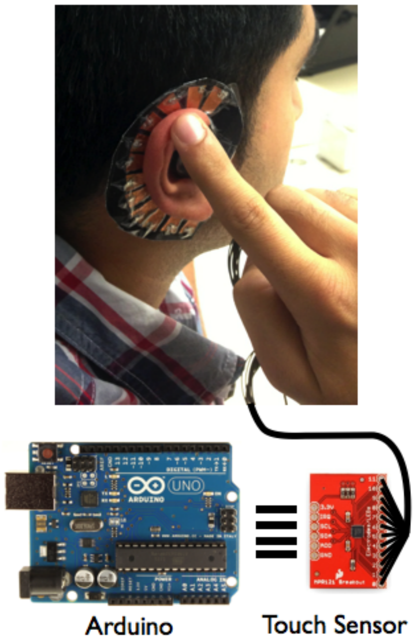


Figure 4.7: Hardware apparatus used in the controlled experiment.

4.3.1.2 Experimental Setup and Methodology

The tasks consisted of simple touch tasks, where the participants had to map a visualized 1-D region-based user interface (comparable to a selection menu) to their ear arc and touch the highlighted area (see Figure 4.8). Technically, the areas were mapped to the corresponding electrodes on the EarPut prototype. The beginning of the ear helix is mapped to the first electrode and the ear lobe to the last electrode.

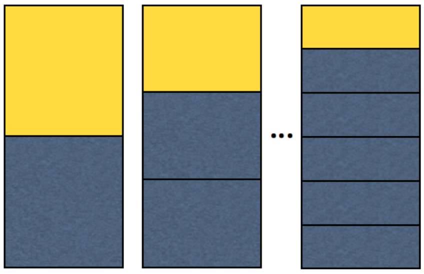


Figure 4.8: Region-based user interfaces used in the experiment. The UIs were subdivided into two to six areas, requiring the participants to touch the highlighted areas.

The experiment was subdivided into two parts: a learning phase and the actual experimental phase. During the learning phase, the on-screen interface provided visual feedback for the touched area. Thus, the participants could familiarize themselves with the functionality of the prototype. During the experimental phase, the on-screen interface only showed the highlighted area and did not provide any visual feedback with respect to the participant's performance. The system advanced to the next target after each touch, regardless of whether the participant had successfully touched the area.

We chose a within-subject design with 27 participants (males, 23; females, 4; average age, 27 years). The independent variable was the amount of areas, considering region-based interfaces with two to six different equally sized areas. The dependent variable was the success rate of a user touching the highlighted region on her ear arc. During the experiment, the participants were seated. After each task, we asked the participants to touch the table to prevent relative positioning of the touches. Each single-user session lasted approximately 15 minutes.

4.3.1.3 Results

For each region-based interface, the participants had to touch each individual area three times (e.g., the interface with two regions resulted in 2 x 3 touch tasks, three per area). The order of the target areas was fully counterbalanced. Overall, each participant had to complete 60 [1] touch tasks leading to 60 x 27 = 1620 data points in total for the experimental phase. We did not collect any data during the learning phase.

The average touch effectiveness of the individual touch areas for each region-based user interface is visualized in Figure 4.10. In the case of two areas, the participants touched both areas equally well. In the other conditions, the upper and the lower parts of the ear arc were touched more effectively than the parts in the middle. Across all conditions,

[1] 60 = 3 repetitions x (2 + 3 + 4 + 5 + 6 areas)

the average effectiveness for touching the ear lobe was above 90% and at least 81% for the upper part of the ear helix. Figure 4.9 shows the average effectiveness of targeting areas per region-based user interface. The effectiveness decreased monotonically over all conditions. The average effectiveness is above 80% for region-based interfaces with up to four areas and decreases to 64% for five and 58% for 6 areas, respectively. ANOVA tests with Bonferroni post-hoc tests revealed that all differences but the one between three and four areas are statistically significant ($p < .001$). The decrease in effectiveness is in line with qualitative findings from the semi-structured interviews. The participants stated that it was hard to precisely distinguish between more than four areas. Beyond region-based interfaces, the participants envisioned more advanced interactions, such as gestures, multi-touch, or grasping. We incorporate the more advanced interaction techniques in our design space and we then show in our application section (subsection 4.3.4) how to leverage them for ear-based interaction.

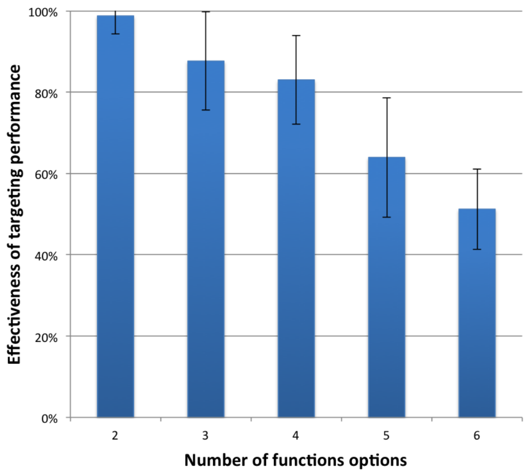


Figure 4.9: The average effectiveness of targeting areas per region-based user interface.

4.3.1.4 Discussion and Summary

The results from the experiment show that users can touch certain areas of their ear arc precisely and effectively, such as the ear lobe (>90%). For an odd total amount of areas, the middle part of the ear arc is more difficult to touch precisely. Thus, both upper and lower parts of the ear arc afford more fine-grained interaction than the middle part (see Figure 4.10(3), Figure 4.10(5)). Consequently, interface elements should not be distributed equidistantly alongside the ear arc, but instead elements placed at the middle part of the arc should be larger than those at the ends.

This finding is also interesting for continuous interactions, such as sliding along the ear arc. To give a simple example: the results suggest that gestures starting at the outer parts

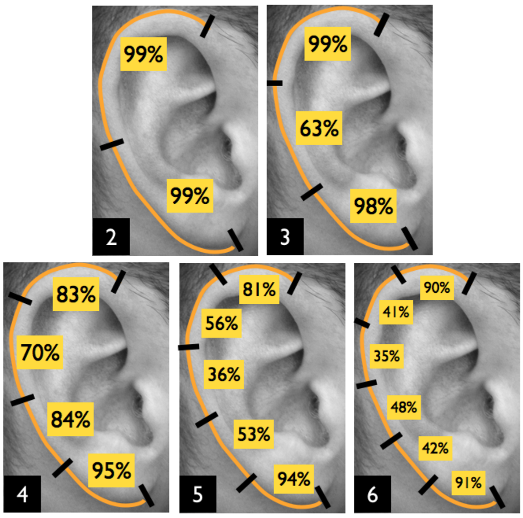


Figure 4.10: The average touch effectiveness for each individual area per region-based user interface.

of the arc (either lobe or upper helix) toward the middle tend to be less error prone than gestures starting in the middle.

Furthermore, our results provide evidence that users can distinguish up to four salient regions on their ear arc effectively (>83%). We envision this to be leveraged as region-based shortcuts, as well as for multi-touch interactions on multiple areas for future ear-based interfaces.

In the interviews, the participants repeatedly suggested to use a variety of other atomic interaction primitives, besides single touch, for ear-based interaction. We transcribed the interviews, selected salient mentions of primitives and analyzed them using an open coding approach. This enabled us to get a first, systematic understanding of the interaction design space, which we present in the following.

The summary of our study results is presented in Table 4.2.

4.3.2 Design Space for Ear-Based Interaction

We transcribed the qualitative data and analyzed it using an open coding approach [Strauss and Corbin, 2008]. The results are three major categories for ear-based interaction. First, considers the input *on* the ear (Ear Centric Interaction, further called *Touch Interaction*). Second, allows for interaction *with* the ear, for example, bending or pulling the ear (also ear-centric interaction, further called *Grasp Interaction*). Last, focuses on interactions around the ear (ear proximity-based interaction, further called *Mid-air Ges-*

Summarized Evaluation Results:
– Our results provide evidence that users can distinguish up to four salient regions on their ear arc effectively (>83%). Users can touch certain areas of their ear arc precisely and effectively, such as the ear lobe (>90%).
– For an odd total amount of areas, the middle part of the ear arc is more difficult to touch precisely. Consequently, interface elements should not be distributed equidistantly alongside the ear arc, but instead, elements placed at the middle part of the arc should be larger than those at the ends.
– Beyond region-based interfaces, the participants envisioned more advanced interactions, such as gestures, multi-touch, or grasping. We incorporate the more advanced interaction techniques in our design space, and we then show in our application section (subsection 4.3.4) how to leverage them for ear-based interaction.

Table 4.2: Summary of preliminary EarPut study.

tures). Within these, various interaction primitives can be used to facilitate ear-based interactions. This in our case allows users to individually interact with sound during fluid collaboration.

Touch Interaction (Figure 4.11 left): The whole ear arc can be used for *single touch* and *multi-touch* input, enabling the user to perform discrete and continuous gestures similar to those found in traditional touch surfaces, for example, a one-finger *sliding* gesture or a two-finger *pinch*.

Grasp Interaction (Figure 4.11 middle): Grasp interactions comprise *bending* or *pulling* the earlobe or the upper helix, as well as *covering* the whole ear. The deformation of the ear is sensed and can be used as both continuous and discrete input.

Mid-Air Gestures (Figure 4.11 right): Mid-air gestures close to the ear can be sensed and used as continuous or discrete input, similar to Metzger *et al.* [2004]. *Hovering* with the hand above the ear can be sensed for distance-based interactions. Then *swiping* the hand near the ear allows for directional interactions.

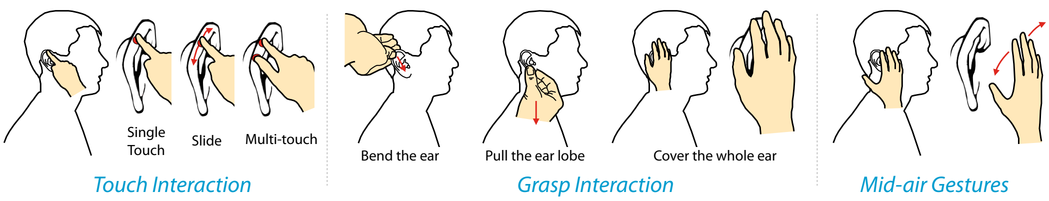


Figure 4.11: Input design space for ear-based interaction.

4.3.3 EarPut: Individual Sound Control with Ear-Based Interaction

The results from the preliminary controlled experiment underline the general feasibility of ear-based interactions. Building upon these results, we developed EarPut: a smart earpiece that augments accessories that are placed or worn behind the ear. EarPut instruments the ear as an interactive surface to enable eyes-free, mobile interaction with the ear (see Figure 4.12). The general objective was to develop a hardware prototype that is unobtrusive and lightweight and that requires little setup (see Figure 4.6). Since EarPut primarily focuses on input, we envision it as a companion device that piggybacks onto existing feedback mechanisms, for example, to wirelessly trigger auditory or vibrotactile feedback through actuators of a smart phone.

4.3.3.1 Hardware Design

To achieve an appropriate hardware footprint, we developed a custom printed circuit board (see Figure 4.13). The main components of the board are an MPR121 Capacitive Touch Sensor [Semiconductor, 2010] used for recognizing touch events, a Bluegiga BLE 113 for Bluetooth communication [Bluegiga, 2013], and an Atmel ATmega1284P microcontroller [Atmel, 2009], which coordinates the measurement and the communication. The EarPut device is powered with a lithium-polymer battery 3.7 V at 110 mAh allowing EarPut to stay functional for over 2 days.

Figure 4.12: The ear hook of devices that are placed or worn behind the ear can be augmented to instrument the human ear as an interactive surface.

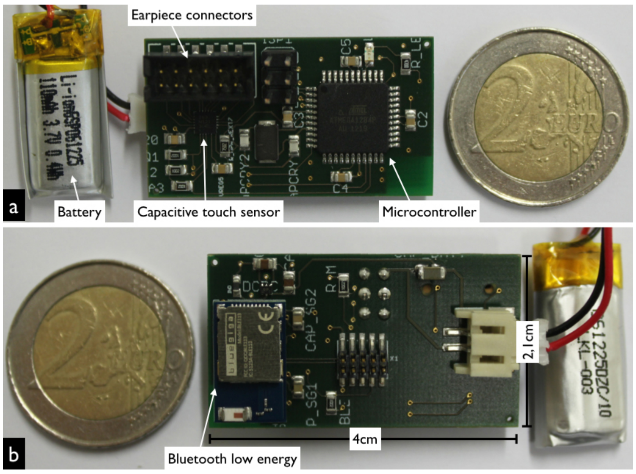


Figure 4.13: EarPut circuit board (a) front and (b) back

We use a similar approach as in the controlled experiment to identify touch-based interactions with the ear. Hereby, the same electrodes as in the experimental apparatus are placed onto an arc-shaped cardboard area. The arc is then used to augment devices worn on or behind the ear, for example, the hook of behind-the-ear earphones or glasses. This enables interaction along the entire ear arc. The electrodes on the ear arc are connected to the circuit board through a ribbon cable.

In comparison with previously presented apparatus used in the preliminary study (see subsubsection 4.3.1.1), EarPut is wireless and can easily be paired with mobile devices via Bluetooth. Furthermore, it has a very compact form factor, is lightweight, and has a flexible adjustable arc-shaped wired cardboard that allows for augmentation of different ear forms.

4.3.3.2 Limitations

The current hardware implementation leverages capacitive sensing. Thus, it can only be used for touch-based interaction if not covered by, for example, objects worn on the head such as hats or caps. A possible solution to this could be to implement the touch input with pressure sensors permitting to still track the touch even when the ear is covered.

4.3.4 Applications

In order to allow on-body interaction, hardware parts needs to be placed on top of the user's body to sense the input. The human ear is particularly suitable for placing this hardware because it allows users to interact with it eyes free, single handed, and bimanual. This is particularly beneficial when users want to have their hands free while being ready for interaction (e.g., "on-the-move" or "slow-move" at their office or at home). Hereby, three different classes of applications exist, where a user is on the move in a computerized environment; however, at the same time, the user expects to be ready for interactions.

Sound control: Our first application allows to control sound by controlling a music player individually with ear-based gestures.

Device remote control: EarPut permits remote controlling of smart object in a computerized environment, for example, light sources or basic control of a TV.

Information control: Information can be placed along the ear arc and accessed via touching the corresponding part of the ear (e.g., a game for memorability).

For each application class, we have implemented one sample application that combined interaction primitives outlined in the design space to compound interactions. In the following, after one another, we present these sample applications.

4.3.4.1 Sounds Control: Music Player

EarPut is highly suitable for controlling sound individually by interacting on the ear. Individual interaction is an important part of fluid collaboration. In the following, we exemplary show how to leverage the interaction primitives to design a music player application. We implemented the music player interface prototypical for the current EarPut device that connects to an Android phone, controlling the stock Android 4.3 media player application.

Basic Navigation

Simple navigation tasks in a media player comprise *play/pause*, navigation to the *next* or *previous* track, and adjustment of the *playback speed* (i.e., fast-forward/rewind). We map these tasks to touch interactions as shown in Figure 4.14a. The ear is subdivided into three regions. A single touch onto the middle region corresponds to play/pause. Tapping the upper or lower regions lets the user navigate within the playlist to the next or previous track. The playback speed can be adjusted by multi-touch gestures in two steps: First, the seeking mode is enabled by tapping the upper and lower region simultaneously (see Figure 4.14b). Second, by holding one of the two touches, the user controls the seeking direction. The user fast-forwards by releasing the lower tap and holding the upper one (see Figure 4.14b top) or rewinds by releasing the upper tap and holding the lower one, respectively (see Figure 4.14b bottom).

Volume Control

Adjusting the playback volume maps naturally to the following sliding interactions alongside the ear arc; sliding from the ear lobe toward the upper ear helix translates to *increasing* the volume; sliding from the ear helix downwards toward the ear lobe translates to *decreasing* the volume (see Figure 4.14c).

Quick Access to Shortcuts

As a more advanced task, we envision particular regions on the ear to serve as shortcuts to previously defined playlists. In line with the findings from our experiment, we subdivide the ear arc into four salient regions (see Figure 4.14d). A single touch onto one of the regions then switches to the corresponding playlist and starts playback.

We employ a cover gesture to allow for an easy mode switch between basic navigation tasks and shortcut access. This is necessary since both interactions employ region-based

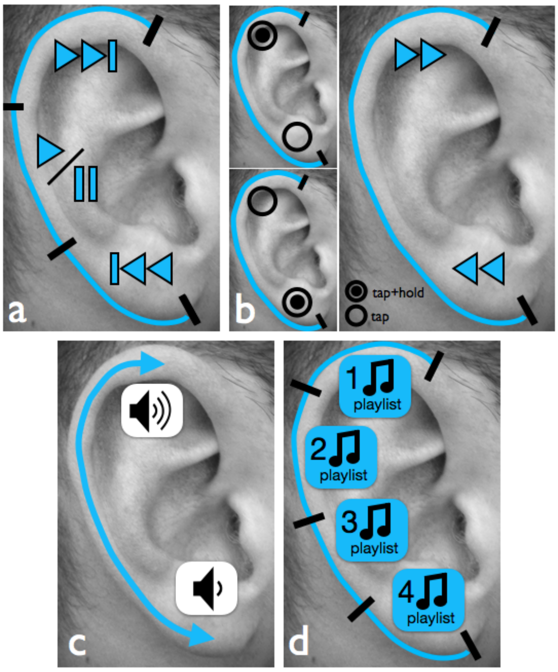


Figure 4.14: Interaction primitives are mapped to design a music player appliance: (a) single touch, (b) multi-touch, (c) slide gestures, and (d) grasp and single touch interactions.

touch interaction. By covering the whole ear, the user switches between the two modes. The current mode is then indicated through auditory feedback.

Similarly, the user could map other interaction primitives such as bending/pulling the ear or performing mid-air gestures to custom tasks individually.

4.3.4.2 Controlling Smart Objects (e.g., Home Appliances)

In addition to the sound control, we envision EarPut to be particularly helpful for controlling home appliances as an omnipresent and eyes-free input device. The next application examples, however, do not contribute to the overarching concept of this thesis, namely, fluid collaboration. In the following, we first show how interaction primitives from our design space could be mapped onto the ear, for example, to select and switch between home appliances. Second, we present an EarPut interface for two application scenarios at home: controlling multiple light sources and a TV remote control.

Select and Switch Control between Home Appliances

A cover gesture wakes up the EarPut in the *home appliances selection mode*. The user is then able to select up to four different home appliances on their ear arc by a single tap on the corresponding region (see Figure 4.15a). After selecting an appliance, a grasp gesture can bring the user back to the home appliances selection mode.

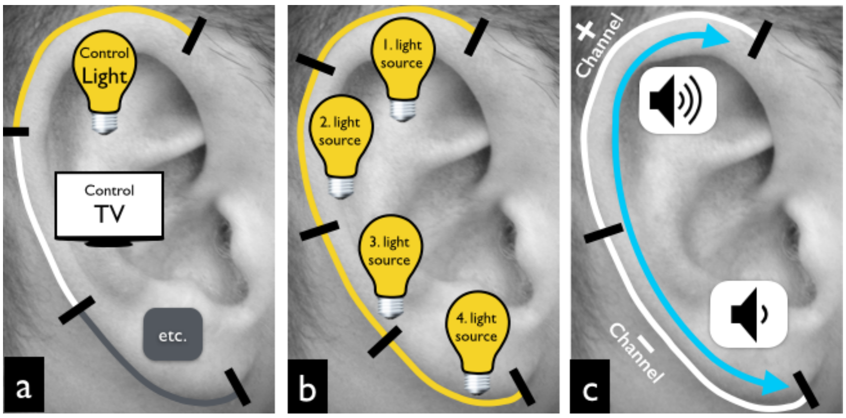


Figure 4.15: Interaction primitives are mapped to control home appliance: (a) a user can select home appliances by single touch on a specific region on the ear arc, (b) single or multi-touch can switch on or off a light source, (c) single touch controls the TV channels and slide gesture controls the volume of the TV.

Control Multiple Light Sources

The user can control up to four light sources. A single or multi-touch on one or more regions triggers the corresponding light sources (see Figure 4.15b).

Simple TV Remote Control

Probably the most frequently used functions of a TV remote are are switching channels and changing volume. EarPut could provide this functionality with a two-region interface to switch channels (see Figure 4.15c) and a slide gesture to control the volume.

4.3.4.3 Information Control: Simon Says Game

Another example application scenario for EarPut is a *concentration game* inspired by the “Simon Says” game [Strommen, 1973]. This application was implemented for a Nexus 4 mobile phone.

In Simon Says, a sequence of actions is proposed to the user visually or via audio and the user has to memorize and repeat the sequence. An action is associated with pressing one of four randomly selected buttons. If the user was successful, another randomly selected button is added to the previous sequence. In the next step, the user has to remember the previous actions, as well as the newly added action and execute them step by step. Consequently, the sequence becomes longer each turn and the game puts the working memory of the player to the test.

For playing the game using EarPut, four different regions on the ear arc are mapped to four different buttons (see Figure 4.16; visual interface only shown for example). When the game starts, the sequence to remember is read to the user through auditory feedback. The user has to repeat the sequence by pressing the corresponding area on the ear arc.

Previous presented EarPut concept allowed users to control sound privately on their ear arc. Next, single and multi-user spatial interaction concepts are presented that allow for focusing on parallel playing sound sources.

4.4 Single-User Spatial Interaction Concepts for Parallel Sound Sources

CoPaperVideo presented in chapter 3 allows spatial interaction with videos on multiple paper-like displays. Since multiple videos can be laid out in physical space and played

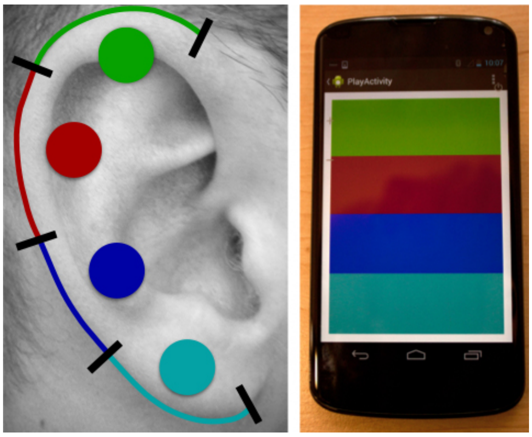


Figure 4.16: Users touches are mapped to design a Simon Says game. Mapping of the ear (left) to the mobile device (right).

back simultaneously, multiple audio sources can be active at the same time. This produces the well-known cocktail party effect [Arons, 1992], which might make it difficult to perceive information conveyed on the audio channel. In this section, we first present interaction concepts that allow a single user to focus on parallel playing sound sources and manipulate them in a direct way. Then we present the technical realization of our system. Then a user study is presented, focusing on evaluating spatial sound concepts in a single-user setting.

4.4.1 Sound Interaction Concepts

The standard case with multiple display devices is that each display has a built-in speaker to generate the audio of the video that is displayed on this device. This makes sure that both visual contents and audio track of one video are located at the same position in space. The sound is perceived in space at a position relative to the user's position and head orientation. Moving the sound source away from the user reduces its volume slightly. We call this sound concept *real-world behavior* (see Figure 4.17a) and implemented it as our baseline.

Next, we introduce three additional sound concepts that will allow a user to more effectively mentally grasp (focus) on one or multiple sound sources that are located in space, reducing the cocktail party effect.

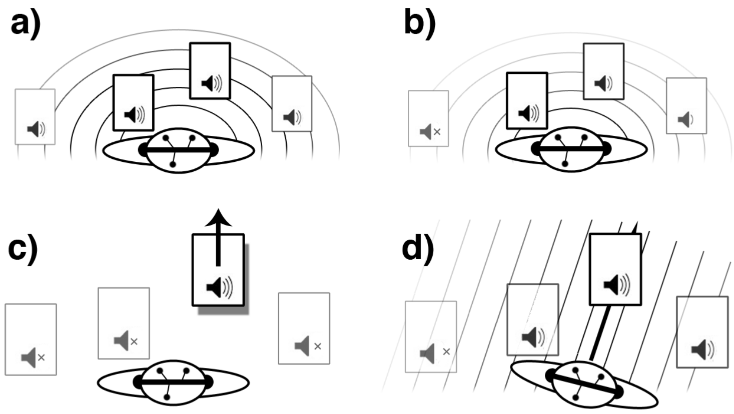


Figure 4.17: Sound concepts a) real-world behavior b) distance-based focusing c) pick-up-based focusing d) orientation-based focusing

4.4.1.1 Distance-Based Focusing

In work with paper documents, it is a well-established practice to focus on documents by placing them directly in front of the user. Documents are brought out of focus by placing them farther away, but still within an arm's reach [Sellen and Harper, 2001]. Inspired by this behavior, we propose a sound concept for focusing on sound sources based on their distance.

Moving a display closer to or more distant from the user increases or decreases its volume (see Figure 4.17b). In contrast to the Real-World Behavior, where distance has only a barely noticeable effect on the volume, the Euclidean distance between the user and the display is mapped inverse exponentially to the volume. As a result, volume can be finely adjusted, somewhat similar to a slider of an audio mixer. Placing the display an arm's length away is distant enough to reduce the volume to zero.

4.4.1.2 Orientation-Based Focusing

When people focus their attention on a person or on an object, they usually look at it. We propose a concept that leverages head orientation of the user for focusing on sound sources.

A virtual line originates from the user's head in the head's orientation. The volume of each display is mapped inverse exponentially to its relative distance from the virtual line. Hence sound originating from displays that are directly within the user's orientation has the highest volume and is located in the center. Sound originating from displays to the left or right side of the line has a lower volume and is located to the left or right of the user. Sound from displays at the extreme outer sides is muted (see Figure 4.17d). By reorienting his head, the user can easily and quickly change his focus to different videos.

4.4.1.3 Pick-Up-Based Focusing

It is common practice to pick up an object to focus on it. Inspired by this behavior, we introduce a further sound concept. Sound from displays that are lying down on the table is set to mute. By lifting up one or multiple displays, the sound of these displays is played back (see Figure 4.17c). While picked up, sound sources expose a Real-World Behavior, being correctly located in space.

4.4.2 Implementation of Virtual 3-D Sound

In order to generate a 3-D perception of sound for a single user, we used the OpenAL Framework [OpenAL, 2010]. This framework uses a well-known approach called Head-Related Transfer Functions (in short HRTFs) to virtually place a sound source in 3-D using headphones (for more detail see subsubsection 4.1.2.3).

The user was equipped with headphones that were augmented with IR markers to track the user's head position and orientation so that the sound sources could be positioned accurately in space. Each playing video on a paper-like display generated a sound relatively positioned to that user's head. With this implementation, we could implement previously presented sound concepts to mentally grasp multiple sound sources simultaneously.

4.4.3 Evaluation of Sound Concepts

In this section, we evaluate and contrast the four sound concepts.

4.4.3.1 Study Design

Our sound evaluation was part of our CoPaperVideo evaluation that was presented in chapter 3 in subsection 3.7.1. We were interested in how a *single* user is able to manipulate and focus on multiple parallel sound sources. Therefore, we added an additional task to the study.

Overview – Study Design	
Method:	Qualitative evaluation
Interest in:	Users' sound practices
Participants:	6 experts (single-user session)
Duration:	avg. 3h
Data gathering:	Semistructured interview, observation and video-taped

Before the last task, the participant was introduced to the four sound concepts. Then he had to perform four subtasks, each with a different sound concept. The order was randomized. In each subtask, he was given a set of five videos on five displays and had to explore and decide which of the videos he liked.

4.4.3.2 Data Gathering and Analysis

For data gathering, we used semi-structured interviews (at the end of a task and after the whole session) and observation. The entire session was videotaped. Interviews and observations were transcribed and analyzed using an open coding approach [Strauss and Corbin, 2008].

4.4.3.3 Results and Summary

In the following, we present the results.

Real-World Behavior

The evaluation showed that the sound concept that realizes real-world behavior is not suitable for viewing videos on multiple displays at the same time. Three participants watched the videos one after another. P5 who tried to view the videos in parallel stated: *"It feels better when only one video is playing."* All other sound concepts where judged to be better than this one.

Distance-Based Focusing

Our observations and comments from the users clearly showed that the distance-based focusing concept is much better suited for watching multiple videos. With this technique, five of six participants watched videos in parallel. Four of six participants rated this to be the best of all concepts, since it allowed for the most flexible sound manipulation with a very intuitive mapping. For instance, P2 stated that *"It is easy to manipulate the volume."* The remaining participant did not watch videos in parallel with any of the concepts.

Orientation-Based Focusing

Three of the five participants who watched videos in parallel criticized this concept because of too much noise coming from displays at the outer sides. Two other participants mentioned that this concept is good for only focusing on audio without visual feedback, but not both combined.

Pick-Up-Based Focusing

The pick-up-based focusing had the advantage that many videos can be played back in parallel without generating sound disturbance. Two participants, P1 and P5, started all the videos right at the start: *"I do not miss anything, I can still see everything."* One participant (P6) mentioned: *"I can focus on one video more clearly."* However, two participants (P1, P4) stressed that they *"have just two hands"* so that they can only hold and listen to two videos at a time. Moreover, one of them (P4) feared that holding the display for a long period could be tiring. P4 proposed as an improvement that *"shortly picking up a video could toggle between active and deactivated sound."*

We conclude that with our sound concepts, in contrast to the real-world behavior, it is possible to watch videos in parallel and user's are able to explicitly and easily change the sound focus. The results show that distance-based focusing, preferred for its high flexibility and intuitiveness, was the best technique. Pick-up-based focusing also has its strengths in situations where users focus only on one or two videos at a time from a set of many videos that are simultaneously played back. A video installation at an exhibition booth is one example.

Table 4.3 summarizes the evaluation results. Our results, however, only focus on single-user interaction. In the following, we present a follow-up study that evaluated the same interaction concepts with multiple users.

4.5 Multi-User Spatial Interaction Concepts for Parallel Sound Sources

Previously presented single-user spatial sound control covers only a part of our goal to support the full spectrum of fluid collaboration (see Definition 7) that allows each user to work individually (perceive individual sound) or in a group of people (perceive group sound) while allowing for flexible transition between both. In the following, we present an iterative design process of collaborative sound concepts that allow users to focus on multiple audio sources individually or in a group of people.

In this section, we first introduce our technical realization of the collaborative virtual 3-D sound. Then we present our preliminary evaluation where we observed how multiple users are perceiving sound with previously presented single-user sound concepts. Results of this evaluation informed our design of the collaborative sound concepts presented

Summarized Results:

Real-World Behavior

✘ This technique is not suitable for listening to parallel sound sources.

Distance-Based Focusing

✔ Highly suitable for focusing on parallel sound sources. five of six participants watching videos in parallel.

✔ This technique allowed the most flexible sound manipulation.

Orientation-Based Focusing

✔ This technique was positively perceived, when users focused only on audio without visual feedback.

✘ For some participants, while focusing on one particular video, the sound from the outer sides was disturbing.

Pick-Up-Based Focusing

✔ Multiple videos can be play backed simultaneously without generating sound disturbance.

✘ Only two displays can be hold up at the same time, forcing the user only to focus on two sound sources at the same time.

Table 4.3: Summarized results for the qualitative evaluation focusing on single-user sound focusing techniques.

next. Lastly, we present the second multi-user evaluation that studies multi-user sound practices with the collaborative sound concepts.

4.5.1 Implementation: Multi-User Extension for Virtual 3-D Sound

The previously presented implementation of virtual 3-D sound provided support *only* for a single user (see subsection 4.4.2). In order to support multiple users to hear virtual 3-D sound individually or collaboratively, previous implementation has been enhanced.

Each user wears their own cable-free headphones. Hereby each headphone has its own radio station that is connected through a Gigaport HD Hub to the pc (see Figure 4.18). Each user's head was instantiated in the OpenAL Framework as a point of reference for a sound source. Users' headphones were augmented with IR markers to track the user's head position and orientation so that the sound sources could be positioned accurately in space for each user.

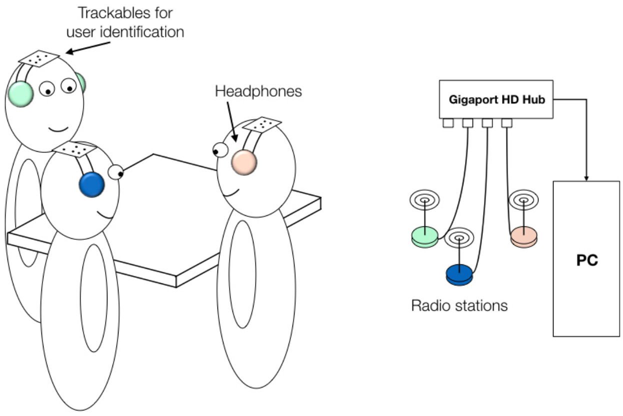


Figure 4.18: Different physical components to support each user with their own sound.

For the second design iteration of the sound concepts, we used nearly the same implementation. The implementation was enhanced with personalized touch input. Personalized touch was implemented to allow touch-based focusing technique. Technical details to realize personalized touch was discussed in chapter 3 in subsubsection 3.6.2.3.

4.5.2 Preliminary Multi-User Evaluation

In order to asses how well *multiple* users can focus on parallel playing sound sources, we have conducted a preliminary user study focusing on a multi-user setting. For this multi-user study, we used the same interaction concepts presented previously for a single user (see subsection 4.4.1). In the following, we report on study design and results.

4.5.2.1 Study Design

Overview – Study Design	
Method:	Qualitative study
Interest in:	Multi-user sound practices
Participants:	9 experts (3 groups of 3)
Duration:	avg. 2h
Data gathering:	Semistructured interview, observation and video-taped

Our multi-user sound study was part of CoPaperVideo multi-user study presented in chapter 3 in subsection 3.7.2. In the following, we only shortly describe the study design for more detail, please refer to the previously mentioned section. The study was conducted with three groups of three people (males, 5; females, 4; average age, 25 years). The study followed a within-subject design with three tasks, one task focusing on the sound concepts and others on spatial interaction concepts with videos. In order to minimize learning effects, all sound concepts for the sound tasks were randomized. We did not measure time during the study. Participants decided themselves, when a task was finished. The average study duration was 2 hours 20 minutes with an in-between break of 30 minutes.

We had a single task that gave us an insight in how well previously presented single-user sound concepts will perform in a collaborative environment when multiple users try to work with videos on multiple paper-like displays. In order to address this question, we asked a group of users: *"Imagine you are planning a party and you are responsible for the music at the party. Please select out of 10 music videos your playlist for the evening. The playlist should have at least 3 music videos."*. This task was performed individually with all four different sound concepts.

After fulfilling the sub-tasks with all the four sound concepts, the participants decided their favorite concept. This sound concept was then used in the other task of the study.

4.5.2.2 Data Gathering and Analysis

As in the previous study, we used semistructured interviews and observation for data gathering. Videos were recorded during the entire session. Interviews and observations were transcribed and analyzed using an open coding approach [Strauss and Corbin, 2008].

4.5.2.3 Results

In the following, we present our study results ordered by the sound concepts.

Real-World Behavior

Participants of groups 1 and 2 agreed after a short time of using the system only to play back one video at the same time. Participants of group 3 first played back videos in parallel then also came to the same decision. When multiple videos were played simultaneously, participants had problems to decide which video were their own *"What video am I listing to?"* (P2).

Distance-Based Focusing

While using this sound concepts, all participants played videos in parallel. However, videos from other participants on a distance still can be heard quietly. This was positively noted by P2 and P7, saying *"Sometimes it was nice to listen in others activity."* (P2). P5, however, found it disturbing, saying *"I think muting the sound from a certain distance would be helpful."* During the task, participants passed videos on to others. Toward the end of the task in group 1 and 3, participants closely gathered together to play selected videos again. Passing content and closely gathering gave indications that this sound concept stimulated collaboration.

Orientation-Based Focusing

During orientation-based focusing, all participants played back videos simultaneously. To hear videos played back by other participants by orienting the head towards them was positively perceived: *"It's really fun to choose songs together, because you can always hear a bit from the others."* (P8). Participants also reported to have an easy control of the video sound: *"You have a great control."* (P3). However, if more than one relevant videos was in the field of vision, users had difficulties to focus: *"It is difficult to watch the videos in a line."* (P4).

Pick-Up-Based Focusing

Participants had difficulties while playing multiple videos simultaneously. Group 2 agreed to play only one video at the same time. Group 3, however, played videos in parallel. Towards the end of the task, group 3 combined their result together by successively picking up their displays.

4.5.2.4 Discussion and Summary

Two of the four sound concepts such as real-world behavior and pick-up-based focusing only marginally allowed users to work with videos and their sounds collaboratively. Participants stated that they have problems to concentrate on their content, because other users playing sounds disturb them. Furthermore assigning a sound to a video was complicated. These two sound concepts are beneficial for single-user setting, however, cannot be used in a collaborative environment. Instead of picking a display up to hear the sound collaboratively, participants of group 1 and 2 expressed the idea to activate sound via touching the display. When multiple users are touching a display, all of them can hear the sound. This also overcomes the limitation of two active displays during the pick-up-based focusing technique. This technique is similar to a previously presented work by Morris *et al.* [2004] and will be in the following presented as a touch-based focusing concept.

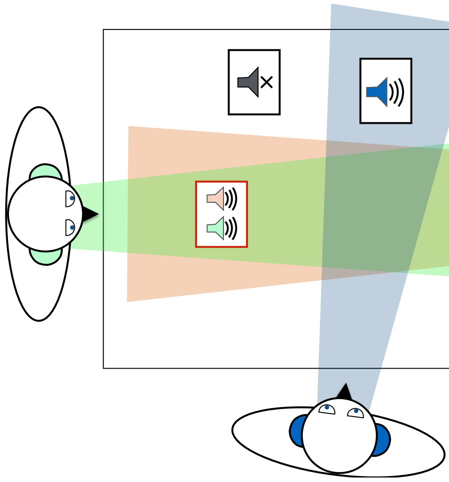


Figure 4.19: Limitation of the orientation-based focusing technique in the collaborative setting. Users cannot distinguish multiple sound sources from a certain direction.

Sound concepts such as distance-based focusing and orientation-based focusing were better received by the users because they allowed each user privately and in a group to focus on different videos. Distance-based focusing allows to focus on sound source near a user while playing sound sources on a distance exponentially quieter. For some of the participants, however, the quite sound was still disturbing. They stated that they wish to have different territories, sound not in their territory should not be heard by them. Different territories are also known from space management; however, they have never been used to control sound. In the next sound concept iteration, we will present a collaborative sound concept that is based on the private and group territoriality introduced by Scott *et al.* [2004].

Participants preferred orientation-based focusing concept, because of its explicit sound control. The only limitation mentioned by participants was that they could not easily distinguish multiple sound sources that were in front of them at the same direction (see Figure 4.19). The assumption hereby was that all users hear all the sound sources in their orientation direction. This situation, however, often occurs during collaborative work, for example, when two users are standing on opposite sides to each other. We address this problem in the following by allowing each user to select between sound sources in a certain direction by tilting the head up and down. We call this technique spot-based focusing and explain this in detail in the next section.

Table 4.4 summarizes the evaluation results. Results of this evaluation informed our design of collaborative sound concepts presented in the following section.

Summarized Results:
Real-World Behavior
✘ Similar as in the single-user setting, this concept marginally allowed users to work with videos and their sounds collaboratively. ↪ This concept is no longer pursued.
Distance-Based Focusing
✔ This concept was positively perceived by the users for focusing on parallel sound sources individually and collaboratively.
✘ This concept allows users to focus on sound sources that are near a user. The volume of the sound on a particular distance to the user are played exponentially quieter. For some of the participants, however, the quite sound was still disturbing. ↪ Different territories are introduced to allow users to focus only on sound sources that are in their private or the public territory.
Orientation-Based Focusing
✔ This technique was positively perceived, because of its explicit sound control.
✘ Users could not easily distinguish multiple sound sources that where in front of them at the same direction. ↪ We address this problem in the following by allowing each user to focus on sound sources only in a certain spot they are looking at. Users can position the spot by rotating and tilting their head.
Pick-Up-Based Focusing
✘ Participants had problems to concentrate on their own content, because other user's playing sound disturbed them. ↪ Instead of picking a display up to hear the sound collaboratively and limit the amount of displays that can be picked up to two, participants of group 1 and 2 expressed the idea of activating sound with touching the display.

Table 4.4: Summarized results for preliminary multi-user study.

4.5.3 Collaborative Sound Concepts

In the following, we present collaborative sound concepts such as spot-based focusing, territory-based focusing, and touch-based focusing. These techniques have been derived in an iterative design process. Hereby the previous multi-user evaluation was used as input to design the following techniques. These concepts support fluid collaboration with sound by allowing multiple users to focus on multiple sound sources individually and collaboratively while supporting a flexible transition between them.

4.5.3.1 Spot-Based Focusing

Every day, people focus on physical objects. Hereby by nature, we only focus on a certain area. Inspired by the real world, we developed the *spot-based focusing* technique. In the center of the spot, the user can hear the maximum volume of the sound. Sound volume decrease exponentially around the spot (GL7 - Individual Sound, see subsection 4.1.3). The user can move the center of the spot forward and backward by tilting the head up and down or turn it left and right (see Figure 4.20) (GL9 - Fluid Sound). Multiple users can hear one sound source when multiple users point their spot onto one sound source (GL8 - Group Sound).

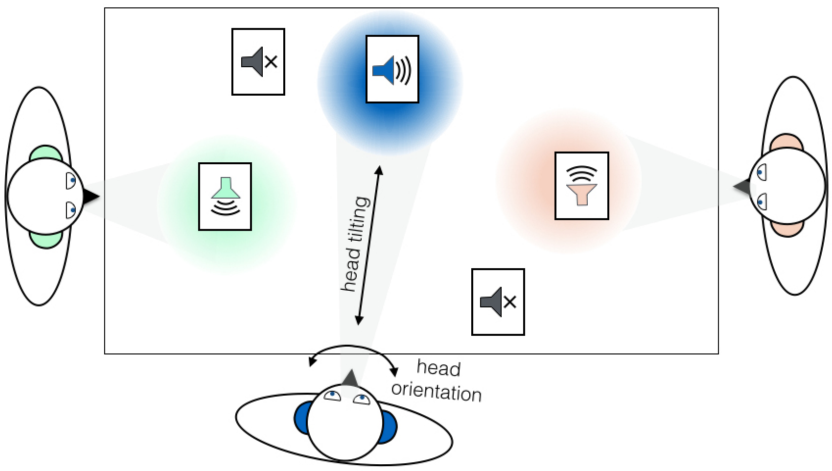


Figure 4.20: Spot-based focusing sound concept provides a spot where a user can privately focus on sounds in a corresponding small area. The user can move the spot by turning or tilting the head, similar to visually focusing on a certain area.

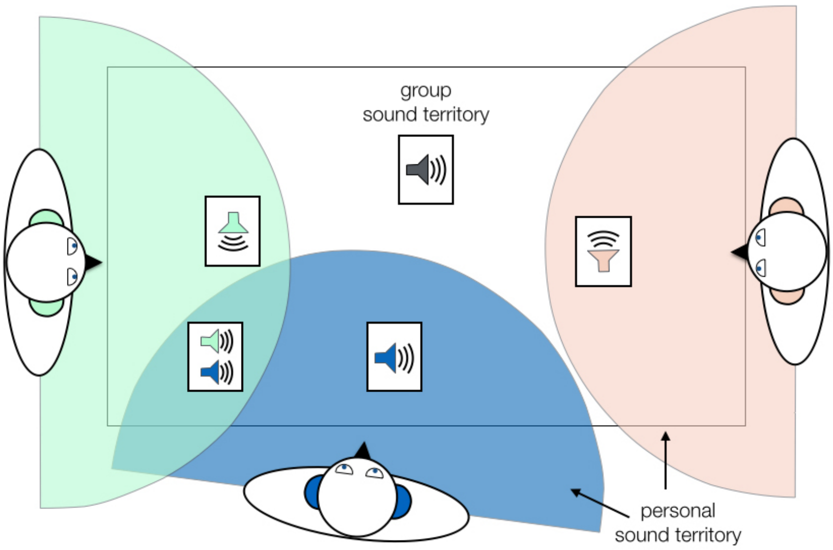


Figure 4.21: Territory-based focusing sound concept allows each user to focus on sound sources inside their private and group territory.

4.5.3.2 Territory-Based Focusing

While working in a group, spatial mappings like private and group territories exist, which have been found by Scott *et al.* [2004]. Inspired by this, we present a territory-based focusing technique. Each user has his own private territory that is generated dynamically around the user's position. Space that is not in one of the user's private territory is considered group territory. Sound sources in private territory can only be heard by the corresponding user (GL7). Sound in group territory can be heard by all users (see Figure 4.21) (GL8). Users can move displays between private and group territory to switch whether the sound can be heard privately or in a group (GL9). A dynamic color frame on the display indicated the corresponding territory. When displays were inside of the territory of multiple users, multiple colored frames were visualized on the display.

The idea of using territories for a specific output (in our case sound) is not new. Dragicevic and Shi [2009] have introduced document orientation techniques based on different territories. When one user moved the document from one territory to another, the document reorients automatically towards the user. However, nobody has investigated territories in combination with sound.

4.5.3.3 Touch-Based Focusing

The previous user study has indicated that pick-up-based focusing is not really useful for a group of people. Users commented that touching a display to control whether a sound

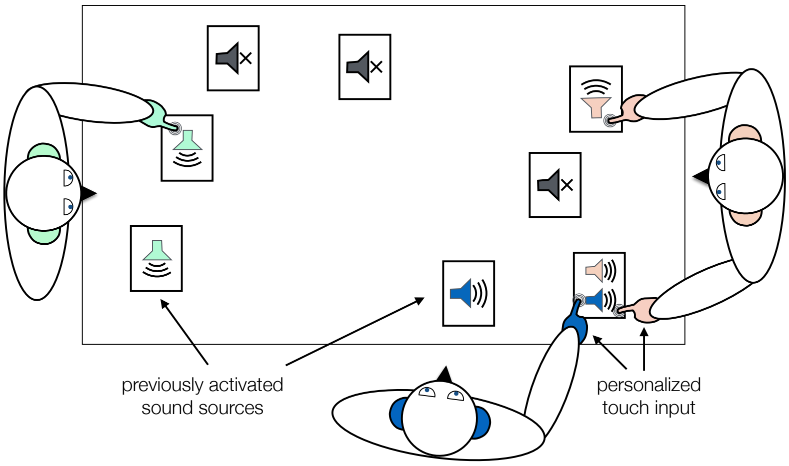


Figure 4.22: Touch-based focusing sound concept allows each user to focus on sound sources by touching the tangible display. Multiple sound sources can also be heard by touching multiple displays.

is played or not would be a better approach. We have adopted pick-up-based focusing and introduced a similar technique we call *touch-based focusing*. Users can privately play back or pause a sound source by touch and hold onto a display (GL7, GL9). Multiple sound sources can be activated and are then played in 3-D space (see Figure 4.22) (GL8).

This technique was inspired by Morris *et al.* [2004]. Morris *et al.*, however, restricted their users to only one sound source at a time. For Morris *et al.* sound was also not located in 3-D space.

4.5.4 Evaluation of Collaborative Sound Concepts

In order to evaluate our collaborative sound concepts, we have conducted an early user evaluation. In the following, we report our study design, data gathering, and results of the study. In particular, we wanted to understand how multiple users would focus on sound in a mixed-focus collaboration scenario. We also focused on the requirements mentioned in section 4.5.

4.5.4.1 Study Design

This study was conducted with three groups of three participants (males, 4; females, 5; average age, 20 years). All participants were familiar with the medium of video (average time of watching video per week was 12.5 hours). All users were standing during the whole study. The average study duration was 2 hours with an in-between break of 15 minutes.

Overview – Study Design	
Method:	Qualitative study
Interest in:	Multi-user sound practices
Participants:	9 experts (3 groups of 3)
Duration:	avg. 2h
Data gathering:	Semistructured interview, observation, and video-taped

During the study, we compared three presented collaborative sound concepts: territory-based focusing, spot-based focusing, and touch-based focusing. Each sound concepts was introduced before the actual task with the sound concepts. Users were allowed to try the sound concept until they felt comfortable.

Our study consisted of two tasks with within-subject design and was randomized to minimize learning effects. Both tasks focused on evaluating the sound concepts. All three sound concepts were counterbalanced and evaluated separately in subtasks.

Task 1 - Create Playlist For each sound concept, the users received 10 music videos on 10 paper-like displays. All displays were piled together and put in the middle of the table. *"Imagine you are planning a party and you are the responsible person for music. You have 10 music videos to choose from. Please put together a playlist for tonight with the top three music videos."*.

Users had to select their favorite sound source before they started with the next task.

Task 2 - Search and Combine Videos In this task, users had to find videos that had particular actors in it, to combine them to a video collage (combined one video combined of multiple videos). Therefore, 30 trailer videos were provided. Before the task, six of the trailer videos were distributed and placed on the displays. The concrete task was: *"Imagine you want to create a video collage of three actors: Brad Pitt, George Cloony, and Julia Roberts. For this purpose, please seek among the movie trailers provided to you and first sort them by actor. Then combine the corresponding videos to one video that will be then your collage. Hint: You will find more videos when you access the related videos"*. The goal of this task was, that users search and playback videos to once more evaluate their choice of the previous selected sound concept.

During the task, participants did not have any time restrictions and decided together when the task was finished.

4.5.4.2 Data Gathering and Analysis

For data gathering, we used semistructured interviews, observation, and video recordings during the entire session. Interviews and observations were transcribed and analyzed using an open coding approach [Strauss and Corbin, 2008].

4.5.4.3 Results

In the following, we report on our results ordered by the introduced sound interaction techniques to the users.

Spot-Based Focusing

While using spot-based focusing technique, the participants of groups 1 and 3 worked mostly in parallel. Hereby, participants also communicated with each other about the video content: *"What do you think about this video here?"* (P7). The participants in group 2, however, played videos sequentially.

This sound concept was perceived as *"exhausting"* (P2) and *"complicated"* (P9). Another disadvantage is also that a user has to visually focus the video all the time: *"You shortly look away and the video you are watching is muted. [. . .] You have to focus the video all the time"* (P1). This was considered as an important disadvantage specially while communicating with other collaborators. Listening to other collaborator's sound was considered comfortable: *"Listening to other user's sound was easy."* (P8).

Territory-Based Focusing

The participants in all groups worked in parallel. Group 1 worked similar to assembly line workers. Hereby, one participant started to watch a video and after finishing passed it to the next one. This working style, however, dissolves after two participants had to wait for a slower one. Group 3 started in the middle of their task to use the group area to play one video at a time, so that everybody could listen to it.

Some users came close to each other to listen to videos together. P5 commented that placing videos in group area would disturb others. *"I would like to use the middle, but* *as soon as I do it others are complaining." "Put it away"* was said by P5 in exactly this situation.

Touch-Based Focusing

During this subtask, all groups worked in parallel. The participants of group 1 and 2 had different opinions regarding this sound concept. On the one hand , participants wanted that all of the group can hear a video: *"I think it's better if all can listen to a video together [...] We then can quickly come to a conclusion."* (P2). They felt that selecting to listen to a video took longer: *"We had to spent three times as long"* (P4). On the other hand, participants said that enabling and disabling sound explicitly was positive: *"I found it good, because I followed my own speed."* (P6),*"It's better, because everyone can decide for themselves, what he hears."* (P3) .

P8 expresses the idea to incorporate a functionality that allows a user to hear exactly the same sound as another user.

For the second task, groups 1 and 2 chose touch-based focusing, and group 3, territory-based focusing as their favorite sound concept.

4.5.4.4 Discussion and Summary

Our study results show that the collaborative sound concepts support fluid collaboration with sound. In general, more participants worked in parallel. Presented concepts have their advantages and also disadvantages that will be first discussed and then summarized in the following. We also compare user's feedback based on the introduced guidelines in subsection 4.1.3.

The sound concept spot-based focusing was considered easy when a user wants to listen to the sound of other users' videos (GL8). However, having to focus on the video during conversations was considered unnatural. Users also mentioned fatigue issues while using their head movement to focus on sound sources (GL7 partially fulfilled). Switching between individual and group sound was perceived as simple and easy (GL9).

The touch-based and territory-based focusing techniques were generally better perceived by the user than the spot-based focusing technique.

During territory-based focusing technique, users mentioned that displays in the group territory were considered cumbersome and distracting from their own sound (GL7 partially fulfilled). Sharing a sound to a group was easy (GL9); others, however, were then distracted by the sounds (GL8 partially fulfilled).

Users rather preferred an explicitly activation and deactivation of sound via touch (GL7). In this respect, this touch-based focusing technique is a valid approach that can be used to focus on individual sounds in a collaborative environment. With touch-based focusing, however, it is difficult to listen to other users' sound (GL8 partially fulfilled) as well as easily switching between own and other users' sounds (GL9 not fulfilled). Depending on where the different users have their videos, they need to be passed over to the corresponding user.

We summarize the study results in Table 4.5. The table compares the presented sound focusing techniques based on the previously presented guidelines (see subsection 4.1.3).

From the table, we can conclude that touch-based focusing is suitable for individual sound and spot-based focusing is well received for group and fluid sound. Hereby, these different sound concepts could be combined or switched during fluid collaboration. The switching, for example, could be done by mapping the different sound concepts onto the user's ear and using the presented EarPut device in section 4.3.

4.6 Conclusion

In this chapter, we have contributed novel devices and interaction and visualization techniques for supporting a direct way of interaction with sound in a fluid collaboration. Our presented contributions were threefold.

For *individual sound interaction*, we contributed EarPut, a novel interface concept and hardware prototype that instruments the ear as an interactive surface for touch-based interactions. This device allows for unobstrusively augmenting earphones to allow a direct control of individual sound by touching one's own ear. Our evaluation results provide empirical evidence that people are able to distinguish between up to four salient areas on their ear arc. This means that up to four different functions can be easily distinguished while interacting on ear. Based on qualitative findings from post-experiment interviews, we systemically set up a first interaction design space for ear-based interaction and present how the primitive interaction can be combined to design applications (e.g., music player).

Furthermore, we have contributed *single-user sound focusing techniques for spatially located virtual sound sources*, hereby, allowing a user to control the volume of multiple sound sources by spatially moving paper-like displays. Our evaluation results showed that one of our proposed distance-based sound-focusing techniques is the most suitable for focusing on parallel sound sources. In addition, we showed that this technique has the most flexible and easy way of manipulating parallel sound individually.

Summarized Results:

Spot-Based Focusing

✔ Users considered it easy when a user wants to listen to the sound of other users videos.
✔ Switching between individual and group sound was perceived as simple and easy.
✘ However, having to focus on the video during conversations was considered unnatural.
✘ Users mentioned fatique issues while using this concept.

Spot-Based Focusing for Fluid Collaboration with Sound:
◐- Individual Sound
●- Group Sound
●- Fluid Sound

Territory-Based Focusing

✔ Sharing a sound by simply moving the display into the group territory was perceived as simple by the users.
✘ Users mentioned that displays in the group territory where considered cumbersome and distracting from their own sound.

Territory-Based Focusing for Fluid Collaboration with Sound:
◐- Individual Sound
◐- Group Sound
●- Fluid Sound

Touch-Based Focusing

✔ Users preferred this technique for focusing on individual sounds in a collaborative environment.
✘ It is, however, difficult to listen to other user's sound and thereby switching between individual and group sound. Depending on where the different users have their videos, displays need to be passed over to the corresponding user for sound activation.

Touch-Based Focusing for Fluid Collaboration with Sound:
●- Individual Sound
◐- Group Sound
○- Fluid Sound

Table 4.5: Summary of study results and comparison of the proposed collaborative sound concepts.

Lastly, we presented *collaborative sound concepts* for individually and collaboratively focusing on multiple sound sources. Our two-step iterative design process that allowed in the first step to inform our collaborative sound concepts design with a preliminary study of previously mentioned single-user concepts in a multi-user environment. Based on users' feedback in the second step, we could design collaborative sound concepts that support individual and group sound as well as flexible transitioning between them. Our evaluation results showed that our concepts support users during fluid collaboration with sound.

Presented techniques could also be useful to inspire bi-manual ear-based interaction, as our current implementation focuses only on single-handed interactions. Also, we have presented work in the area of colocated collaborative TV viewing [Buchner *et al.*, 2014], where users were using their mobile phones as TV companion devices to view and share content with others in a room while watching TV. Hereby, each device can generate their own sound. Our techniques could be beneficial to this field of research.

5 Conclusion

The presented thesis focuses on computer support for *co-located meetings*. Nowadays in nearly every co-located collaboration, digital information supports users (e.g., making decisions or planning process). The goal of this thesis is to support selection, consumption, manipulation, and production of information that is, at least in part, digitally represented (document, videos, etc.).

In co-located collaborations, when working in a group, users have to constantly switch between individual work and group work, which is also known as *mixed-focus collaboration* [Gutwin and Greenberg, 1998]. Current interactive tabletops and mobile devices available during a meeting setting, demand users to partition the screen surface into dedicated personal and group territories [Scott *et al.*, 2004; Tse *et al.*, 2004], which results in limitations, such as (1) limited interaction space, (2) screen clutter, and (3) workspace interference, particularly when users try to flexibly transition between individual work and group work in mixed-focus collaboration (coined as **fluid collaboration**). In addition, current devices in meeting rooms feature both multiple output modalities, such as visual output and sound output.

This thesis went beyond existing devices and focused on both output modalities in a fluid collaboration setting and contributed novel devices, such as (1) an interactive multi-view tabletop for *surface-based interaction*, (2) multiple spatially aware paper-like displays for *spatial interaction* with videos, and (3) interactive earpiece for ear-based sound interaction as well as multiple paper-like displays for spatial interaction with multiple sound sources. Both sound interaction concepts support direct interaction with sound by *embodied sound interaction*. The proposed devices and interaction concepts effectively support fluid collaboration and allow users a seamless and intuitive way of transitioning between individual work and group work, while maximizing proximity and conserving *close phase social distance* [Hall and Hall, 1969].

In the following, we summarize the main contributions of this thesis according to the main research areas and achievements.

5.1 Summary

5.1.1 Surface-Based Interaction

In order to *support individual work and group work on the very same interactive horizontal surface*, a fundamental understanding and rethinking of the atomic display pixel is required. Currently, each display pixel shows content that is visible to all users and all users can interact with it. In order to support private independent views for individual work in addition to a group view, *each pixel should support different levels of visualization*, ranging from private visualization of private information (e.g., views or elements) for each user to shared common information to all users. This concept of private and group views for each user for each display pixel, in combination with personalized input for each user in order to simultaneously interact on the overlapping private and group views, allowed us to explore *surface-based interaction* for fluid collaboration.

Thereby, we contributed a concept that was named Permulin. Permulin is an integrated set of interaction and visualization techniques for multi-view tabletops to support co-located collaboration across a wide variety of collaborative coupling styles. We built and evaluated a working prototype that (1) provides support both for group work and for individual work, as well as for the transitions in between; (2) contributes sharing and peeking techniques to support mutual awareness and group coordination during phases of individual work; (3) reduces interference during group work on a group view; and (4) directly integrates with conventional multitouch input.

Based on two user studies, we can conclude that Permulin supports fluid collaboration by allowing users to *transition fluidly between loose and tight collaboration*. Our results have shown that participants frequently used Permulin's interaction techniques for dividing and merging views, as well as sharing content to coordinate workspaces. Users also utilized Permulin both highly cooperatively and individually. This is reflected by users occupying significantly larger interaction areas on Permulin than on a tabletop system. It also allows users to perform parallel collaboration, particularly on shared full-screen contents. Furthermore, Permulin provides unique awareness properties: participants were highly aware of each other and their interactions during tightly coupled collaboration, while being able to unobtrusively perform individual work during loosely coupled collaboration.

These qualitative and quantitative results indicate that Permulin is a first step toward effective fluid collaboration on an interactive surface.

5.1.2 Spatial Interaction

In the last years, an emergent display technology has evolved that is flexible, thin, and lightweight and has a similar form factor and affordances as paper documents. These future paper-like displays go beyond paper by providing (1) a *high display refresh rate* that supports visualization of high dynamic digital content, and (2) they can be *spatially aware*, by knowing their own and other displays' locations. In this chapter, we contributed a working prototype of nowadays-still-unavailable paper-like displays. We developed interaction and visualization techniques for browsing and viewing high dynamic digital content–in our case videos. With inspiration from the physical world when working in close physical collaboration with paper documents, we investigated how well-known physical interactions with paper can be transferred to the area of video navigation to allow fluid collaboration by spatially interacting with **multiple spatially aware paper-like displays**.

In terms of *spatial interaction*, we proposed a novel paradigm for users to spatially interact with videos. We introduced CoPaperVideo, a collaborative environment for spatial interaction with videos on multiple spatially aware paper-like displays. Based on the design space for these displays, we introduced a set of interaction techniques that support playback, flexible navigation, and spatial organization of videos for both individual and collaborative use.

We evaluated CoPaperVideo with two user studies, with the focus on single and multi-users' interaction with the working system.

Results from two iterative evaluations shed light on how multiple people use multiple interactive paper-like displays simultaneously and how this affects interaction with video contents.

Results from the *single-user evaluation* showed that users can flexibly organize and structure videos in physical space while generating a good overview of multiple videos. They thereby flexibly attribute three different functional roles to paper-like displays: information source, working display, and information container. We have characterized different mental models and strategies of users ("materializers" vs. "virtualizers") to cope with a restricted number of displays.

Multiuser evaluation results indicate that CoPaperVideo is suitable for collaborative use. CoPaperVideo allows users to work individually as well as in a group while allowing transitioning between coupling styles in a similar way as with paper documents. Furthermore, participants could flexibly exchange information and synchronize their working states in a similar way as they do with paper documents.

5.1.3 Embodied Sound Interaction

Meeting rooms equipped with devices (e.g., tabletops, mobile devices that are brought to a meeting) provide multiple output modalities, such as visual output and sound output. In previous research directions, we focused on supporting the visual output modality and contributed novel single and multiple display surfaces that allowed for fluid collaboration with visual content. In this research direction, coined *embodied sound interaction*, we focused on the sound output modality and contributed novel devices and visualization and interaction techniques that support a direct way of interacting with sound in a fluid collaboration. Thereby we presented body-based and spatial interaction techniques for individual and collaborative volume control of multiple sound sources.

Two of the presented contributions focused on individually controlling multiple sound sources, whereas the third contribution supported a fluid switch between individually and collaboratively hearing multiple sound sources. Both spatial interaction techniques allow a single and multiuser to directly manipulate multiple sound sources by virtually positioning them in 3-D space and physically binding them to paper-like displays.

Based on the reason that sound is perceived though our ears, we presented how the human ear can be used for ear-based input to *directly and privately* control sound. Thereby we contributed a novel device, called *EarPut*. EarPut augments accessories that are placed or worn behind the ear (e.g., behind-the-ear earphones, headsets, or Google Glasses) and thereby unobtrusively instruments the ear as an interactive surface. This allows the user to remotely control and manipulate sound sources by privately touching her own ear.

In a controlled experiment with 27 participants, we assessed both precision and effectiveness of single-touch interactions with EarPut. The results provided empirical evidence that people were able to distinguish up to four salient areas on their ear arc. The results showed that the upper and lower parts of the ear arc afford more precise interaction than the middle part. These findings were particularly interesting for region-based interfaces and suggested that a nonequidistant spacing of interface elements alongside the ear arc is more effective. Based on qualitative findings from postexperiment interviews, we systemically set up a first interaction design space for ear-based interaction. We showed how primitive interactions could be combined to design and implement a variety of ear-based applications.

Sound is perceived omnidirectional, whereas the source of the sound is placed in a 3-D space. We contributed several novel spatial interaction concepts with sound that allow a *single* user to mentally grasp and directly control multiple audio sources simultaneously. Thereby sound sources are virtually placed in 3-D space and permit each user to control the sound volume by physically moving paper-like displays.

The first user study showed that advanced spatial sound concepts effectively support a user in simultaneously viewing multiple videos with sound output. Thereby, results indicated that a spatially located 3-D sound allows a user to focus on parallel sound sources simultaneously. The evaluation results also showed that one of our proposed sound focusing techniques (called distance-based focusing) is the most suitable for focusing on parallel sound sources. In this concept, moving a display closer to or more distant from the user increases or decreases its corresponding sound volume. Volume of displays at a distance beyond the arm's length is muted. In addition, we showed that this technique has the most flexible and easiest way of manipulating parallel sounds individually.

We have contributed a working prototype for spatially controlling 3-D sound in a collaborative environment. We have followed a two-step iterative design process to develop collaborative sound concepts. As a first step, a preliminary study was conducted. Thereby single-user concepts were used in a multi-user environment. This first study informed our design in the second step, where we contributed interaction techniques to control sound, allowing each user to *individually and collaboratively* focus on multiple sound sources with a group of people.

Our evaluation results have shown that a combination of our collaborative sound concepts potentially allows a fluid collaboration with sound, thereby allowing a single user to focus and control private sound sources individually while simultaneously allowing multiple users to focus on a group sound together and easily switching between these two states.

In the following, we present potential future research directions that have been identified based on the previous research done for this thesis.

5.2 Potential Future Research Directions

Only recently technology has started to support users during collaboration, particularly where people often switch between individual work and group work. Focusing on this vein of research, this thesis contributed novel devices and interaction concepts for fluid collaboration. However, it is only a start into the direction where technology supports the user while collaborating with others, and a lot more needs to be done. Each of the three presented research areas in this thesis still has open research questions that can be the focus of future research.

Nowadays, geographically spread projects has become popular. These phenomena can be explained with globalization and the advanced technology that is connecting us. Thereby so-called remote collaboration has become a part of our daily work. In terms of *surface-based interaction*, future research could explore how interactive multi-view tabletops with personalized input and output, such as Permulin, can be used for **remote fluid collaboration**. The advantage of Permulin in remote collaboration could be that group members, whether local or remote, could team up and focus on their individual group collaboration in private views, while other collaborators could continue working on their group tasks with a group view. The rises an unexplored research question such as: *How can remote groups effectively collaborate on remote multi-view tabletops during fluid collaboration?*

Another research direction in the same context of *surface-based interaction*, could explore tangible interaction on multi-view tabletops. Although touch input is an often used input technology, it still lacks haptic and natural tactile feedback [Montagu, 1986] and physical affordances [Terrenghi *et al.*, 2007]. Hence tangible interaction on such tabletops has been the focus of many researchers. Permulin, however, needs personalized tangible input, in order to allow each user to control her own private view. Thereby the following research questions arise: *How will tangible user interfaces that provide personalized tangible input look like? How can personalized tangible interaction effectively be used in collaboration on a tabletop with personalized input and output?*

Tabletops are static and do not provide each user with a private view and a group view for **fluid collaboration in a mobile setting**. A solution could be see-through augmented reality glasses (head-mounted display (HMD) e.g., Meta SpaceGlasses[1] or Vuzix STAR 1200XL[2]). Emerging technology in this field has provided smart glasses that allow for visualization of digital content augmented into the real world. Mobile fluid collaboration could occur when multiple collaborators wearing smart glasses would use the same physical noninteractive space (e.g., a table or a wall) as a shared reference. Smart glasses in combination with 3-D tracking (most of the smart glasses nowadays already feature 3-D tracking) could establish same understanding of the physical space and project shared information to all users. Since the glasses allow for visualizing personal information, users could also work individually. This would allow multiple collaborators with only a single device per user to establish a digital co-located room for fluid collaboration. What is interesting and currently still unexplored would be the following research questions: *How could users use a virtual 3-D space for fluid collaboration? How would fluid collaboration would change in a mobile setting?*

In regard to *spatial interaction*, current technological trends in flexible displays show that highly flexible [Samsung, 2013] and resizable displays [Sony, 2010; Vision, 2011] will be commercially available in the near future. In recent years, researchers have gained a high interest in modern topics such as flexible displays [Cassinelli and Ishikawa, 2005; Dand and Hemsley, 2013; Kildal *et al.*, 2012; Lahey *et al.*, 2011; Samsung, 2013; Schwesig *et al.*, 2004; Steimle *et al.*, 2013; Tarun *et al.*, 2013], resizable displays [Khalilbeigi *et al.*, 2010, 2011, 2012; Steimle and Olberding, 2012], and dynamic shape-changing interfaces [Follmer *et al.*, 2013; Yao *et al.*, 2013].

[1] https://www.spaceglasses.com/

[2] http://www.vuzix.com/UKSITE/augmented-reality/products_star1200xl.html

This opens up novel devices and interaction concepts not only for mobile interaction in general but also for video navigation. In addition, in the last couple of years, new video formats has become popular because of advanced video recording technology, featuring a wider view angle of the cameras, for example, 180° or 360° videos. Current displays, however, lack screen space and proper interaction concepts to navigate such videos. Future flexible, resizable, and shape-changing displays alleviate the restrictions of current fixed-size devices. We believe they will therefore pave the way for more usable and enjoyable interaction techniques for video browsing [Cassinelli and Ishikawa, 2005; Lissermann *et al.*, 2012c] due to their flexible screen size and rich physical interactions [Khalilbeigi *et al.*, 2011]. This situation opens up novel opportunity for research in the area of video navigation with flexible displays and raises the following unexplored research questions: *How can flexible displays (e.g., rollable or foldable displays) allow control and navigation of novel video content, for example, 180° or 360° videos?*

In terms of *embodied sound interaction*, we contributed EarPut, which allowed each user to leverage their own ear as an interactive surface, if we imagine that each user will wear one or even two EarPut devices. It is interesting to explore the following: *How could these devices be interconnected? How could interconnected EarPuts allow a user to access the other users' sound or communicate with them?*

Furthermore, in terms of *embodied sound interaction*, we allowed for direct manipulation of sound in both individual and collaborative ways. The volume of virtually 3-D-positioned sound sources could be changed in a way that we are not used to from the real world. This change enabled developing effective concepts for focusing on multiple sound sources. Another way of looking at this is to say that we have changed the perception of sound in physical space, leading to a more enhanced human perception. We believe a lot more needs to be done in terms of changing human sound perception. One example could be *sound spaces*, which could be adaptable for each user or a group of users. These sound spaces consist of multiple sound sources, located in 3-D sound space around each user. Thereby a user could perceive and haptically feel where in 3-D space the sound is located and coming from and then directly manipulate it. One technical realization of feeling the sound in 3-D space could be AIREAL [Sodhi *et al.*, 2013], which enables users to feel virtual 3-D objects through tactile sensations in free air. This project gives users, however, only a point-by-point feeling of the 3-D space. Sound is a linear 1-D medium, and we can assume that a more linear way of controlling and feeling the sound in 3-D space is crucial. For example, music could be haptically felt, perceived, touched, and manipulated in 3-D space in a completely different way by a single or even multiple users. This would allow a novel way of haptically feeling the sound and would allow for novel ways of interacting with it. This could generate and allow the following research questions: *How can sound be felt in 3-D space to allow a single or multiple users to directly manipulate sound?*

This thesis presented three different research directions, which focused on single and multiple interactive surfaces, and explored how these surfaces allow a single or multiple users to interact with visual and sound outputs during fluid collaboration. This thesis, however, did not combine these contributions into a single system and studied how these different novel devices could work together and how users would interact with a combined system. While it seems promising, it exceeds this thesis and thereby remains for future work.

Bibliography

Mortimer Jerome Adler and Charles Lincoln Van Doren. *How to read a book*, volume Rev. and u. Simon and Schuster, 1972. ISBN 067121280X. URL `http://www.loc.gov/catdir/enhancements/fy0641/72081451-s.html`. (Cited on pages 89 and 103.)

Maneesh Agrawala, Andrew C. Beers, Ian McDowall, Bernd Fröhlich, Mark Bolas, and Pat Hanrahan. The two-user responsive workbench: Support for collaboration through individual views of a shared space. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 327–332, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. ISBN 0-89791-896-7. doi: 10.1145/258734.258875. URL `http://dx.doi.org/10.1145/258734.258875`. (Cited on pages 20, 39 and 49.)

Brian Amento, Will Hill, and Loren Terveen. The sound of one hand: A wrist-mounted bio-acoustic fingertip gesture interface. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 724–725, New York, NY, USA, 2002. ACM. ISBN 1-58113-454-1. doi: 10.1145/506443.506566. URL `http://doi.acm.org/10.1145/506443.506566`. (Cited on page 137.)

Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. Medusa: A proximity-aware multi-touch tabletop. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 337–346, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047240. URL `http://doi.acm.org/10.1145/2047196.2047240`. (Cited on pages 28, 37 and 61.)

Apple. iTunesU. `http://www.apple.com/education/ipad/itunes-u/`, 2007. [Online; accessed 11-Mar-2014]. (Cited on page 106.)

Barry Arons. A review of the cocktail party effect. *JOURNAL OF THE AMERICAN VOICE I/O SOCIETY*, 12:35–50, 1992. (Cited on pages 139 and 160.)

Daniel Ashbrook, Patrick Baudisch, and Sean White. Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2043–2046, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979238. URL `http://doi.acm.org/10.1145/1978942.1979238`. (Cited on page 137.)

Atmel. 8-bit avr microcontroller with 128k bytes in-system programmable flash: Atmega1284p, 2009. (Cited on page 154.)

Thomas Augsten, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch. Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 209–218, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0271-5. doi: 10.1145/1866029.1866064. URL `http://doi.acm.org/10.1145/1866029.1866064`. (Cited on page 34.)

Ronald M. Baecker, Jonathan Grudin, William Buxton, and Saul Greenberg. *Readings in Human-Computer Interaction: Toward the Year 2000.* (Second Edition) Morgan Kaufmann Publishers, Inc., January 1995. ISBN 1558602461. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1558602461`. (Cited on page 5.)

Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1239–1248, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208576. URL `http://doi.acm.org/10.1145/2207676.2208576`. (Cited on page 137.)

Till Ballendat, Nicolai Marquardt, and Saul Greenberg. Proxemic interaction: Designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 121–130, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0399-6. doi: 10.1145/1936652.1936676. URL `http://doi.acm.org/10.1145/1936652.1936676`. (Cited on page 100.)

Jacob T. Biehl and Brian P. Bailey. Improving interfaces for managing applications in multiple-device environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pages 35–42, New York, NY, USA, 2006. ACM. ISBN 1-59593-353-0. doi: 10.1145/1133265.1133273. URL `http://doi.acm.org/10.1145/1133265.1133273`. (Cited on page 22.)

Eric A. Bier and Steven Freeman. Mmm: A user interface architecture for shared editors on a single screen. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*, UIST '91, pages 79–86, New York, NY, USA, 1991. ACM. ISBN 0-89791-451-1. doi: 10.1145/120782.120791. URL `http://doi.acm.org/10.1145/120782.120791`. (Cited on page 30.)

Bluegiga. Ble113: Preliminary data sheet, 2013. (Cited on page 154.)

John Bolton, Kibum Kim, and Roel Vertegaal. A comparison of competitive and cooperative task performance using spherical and flat displays. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, pages 529–538, New York, New York, USA, February 2012. ACM Press. ISBN 9781450310864. doi: 10.1145/2145204.2145286. URL `http://dl.acm.org/citation.cfm?id=2145204.2145286`. (Cited on page 69.)

Guilherme Boreki and Alessandro Zimmer. Hand geometry: A new approach for feature extraction. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, AUTOID '05, pages 149–154, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2475-3. doi: 10.1109/AUTOID.2005.33. URL `http://dx.doi.org/10.1109/AUTOID.2005.33`. (Cited on page 32.)

Karolina Buchner, Roman Lissermann, and Lars Erik Holmquist. Interaction techniques for co-located collaborative tv. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, New York, NY, USA, 2014. ACM. (Cited on page 179.)

Vincent Buil, Gerard Hollemans, and Sander van de Wijdeven. Headphones with touch control. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices &Amp; Services*, MobileHCI '05, pages 377–378, New York, NY, USA, 2005. ACM. ISBN 1-59593-089-2. doi: 10.1145/1085777.1085877. URL `http://doi.acm.org/10.1145/1085777.1085877`. (Cited on page 138.)

Alvaro Cassinelli and Masatoshi Ishikawa. Khronos projector. In *ACM SIGGRAPH 2005 Emerging Technologies*, SIGGRAPH '05, New York, NY, USA, 2005. ACM. doi: 10.1145/1187297.1187308. URL `http://doi.acm.org/10.1145/1187297.1187308`. (Cited on pages 186 and 187.)

Jessica R. Cauchard, Markus Löchtefeld, Pourang Irani, Johannes Schoening, Antonio Krüger, Mike Fraser, and Sriram Subramanian. Visual separation in mobile multi-display environments. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 451–460, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047256. URL `http://doi.acm.org/10.1145/2047196.2047256`. (Cited on page 22.)

Ke-Yu Chen, Kent Lyons, Sean White, and Shwetak Patel. utrack: 3d input using two magnetic sensors. In *Proceedings of the 26th Annual ACM Symposium on User Interface*

Software and Technology, UIST '13, pages 237–244, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502035. URL http://doi.acm.org/10.1145/2501988.2502035. (Cited on page 137.)

Elise Co and Nikita Pashenkov. Emerging display technologies for organic user interfaces. *Commun. ACM*, 51(6):45–47, June 2008. ISSN 0001-0782. doi: 10.1145/1349026.1349036. URL http://doi.acm.org/10.1145/1349026.1349036. (Cited on pages 3, 12 and 87.)

G. Crawford. *Flexible Flat Panel Displays*. Wiley Series in Display Technology. Wiley, 2005. ISBN 9780470870488. URL http://books.google.de/books?id=U9SsFZw-_IMC. (Cited on pages 3, 12 and 87.)

Dhairya Dand and Robert Hemsley. Obake: Interactions on a 2.5d elastic display. In *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13 Adjunct, pages 109–110, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2406-9. doi: 10.1145/2508468.2514734. URL http://doi.acm.org/10.1145/2508468.2514734. (Cited on page 186.)

Chi Tai Dang, Martin Straub, and Elisabeth André. Hand distinction for multi-touch tabletop interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 101–108, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-733-2. doi: 10.1145/1731903.1731925. URL http://doi.acm.org/10.1145/1731903.1731925. (Cited on page 32.)

Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. Palmrc: Imaginary palm-based remote control for eyes-free television interaction. In *Proceedings of the 10th European Conference on Interactive Tv and Video*, EuroiTV '12, pages 27–34, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1107-6. doi: 10.1145/2325616.2325623. URL http://doi.acm.org/10.1145/2325616.2325623. (Cited on page 137.)

Paul Dietz and Darren Leigh. DiamondTouch: a multi-user touch technology. In *Interface*, volume 3, pages 219–226. ACM, 2001. ISBN 158113438X. doi: 10.1145/502348.502389. URL http://portal.acm.org/citation.cfm?id=502389. (Cited on pages 23, 25, 37, 61 and 141.)

Alan Dix, Janet Finlay, Gregory Abowd, and Russell Beale. *Human-computer Interaction.* (Second Edition) Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998. ISBN 0-13-437211-5. (Cited on page 5.)

K. C. Dohse, Thomas Dohse, Jeremiah D. Still, and Derrick J. Parkhurst. Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. In *Proceedings of the First International Conference on Advances in Computer-Human Interaction*, ACHI

'08, pages 297–302, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-0-7695-3086-4. doi: 10.1109/ACHI.2008.11. URL http://dx.doi.org/10.1109/ACHI.2008.11. (Cited on pages 32 and 37.)

Paul Dourish. *Where the Action is: The Foundations of Embodied Interaction*. A Bradford book. Bradford Books, 2004. ISBN 9780262541787. URL http://books.google.de/books?id=DCIy2zxrCqcC. (Cited on pages 5, 13 and 136.)

Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92, pages 107–114, New York, NY, USA, 1992. ACM. ISBN 0-89791-542-9. doi: 10.1145/143457.143468. URL http://doi.acm.org/10.1145/143457.143468. (Cited on page 7.)

Pierre Dragicevic and Yuanchun Shi. Visualizing and manipulating automatic document orientation methods using vector fields. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 65–68, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-733-2. doi: 10.1145/1731903.1731918. URL http://doi.acm.org/10.1145/1731903.1731918. (Cited on page 172.)

Jörg Edelmann, Yvonne Kammerer, Birgit Imhof, Peter Gerjets, and Wolfgang Straßer. Tangoscope: A tangible audio device for tabletop interaction. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part III*, INTERACT'11, pages 427–434, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-23764-5. URL http://dl.acm.org/citation.cfm?id=2042182.2042220. (Cited on pages 141, 142, 145 and 221.)

George W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Commun. ACM*, 36(7):39–49, July 1993. ISSN 0001-0782. doi: 10.1145/159544.159566. URL http://doi.acm.org/10.1145/159544.159566. (Cited on pages 90, 92, 99 and 131.)

George W. Fitzmaurice, Hiroshi Ishii, and William A. S. Buxton. Bricks: Laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 442–449, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-84705-1. doi: 10.1145/223904.223964. URL http://dx.doi.org/10.1145/223904.223964. (Cited on page 91.)

Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. inform: Dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 417–426, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502032. URL http://doi.acm.org/10.1145/2501988.2502032. (Cited on page 186.)

Masaaki Fukumoto and Yoshinobu Tonomura. Whisper: A wristwatch style wearable handset. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 112–119, New York, NY, USA, 1999. ACM. ISBN 0-201-48559-1. doi: 10.1145/302979.303009. URL http://doi.acm.org/10.1145/302979.303009. (Cited on page 138.)

William Gardner. 3d audio and acoustic environment modeling, 1999. (Cited on pages 142, 143 and 221.)

James R. Glass, Timothy J. Hazen, D. Scott Cyphers, Igor Malioutov, David Huynh, and Regina Barzilay. Recent progress in the mit spoken lecture processing project. In *INTERSPEECH*, pages 2553–2556. ISCA, 2007. (Cited on page 94.)

Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 3–12, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0271-5. doi: 10.1145/1866029.1866033. URL http://doi.acm.org/10.1145/1866029.1866033. (Cited on page 137.)

Sean Gustafson, Christian Holz, and Patrick Baudisch. Imaginary phone: Learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 283–292, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047233. URL http://doi.acm.org/10.1145/2047196.2047233. (Cited on page 137.)

Sean G. Gustafson, Bernhard Rabe, and Patrick M. Baudisch. Understanding palm-based imaginary interfaces: The role of visual and tactile cues when browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 889–898, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2466114. URL http://doi.acm.org/10.1145/2470654.2466114. (Cited on page 137.)

Carl Gutwin and Saul Greenberg. Design for individuals, design for groups. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work - CSCW '98*, pages 207–216, New York, New York, USA, November 1998. ACM Press. ISBN 1581130090. doi: 10.1145/289444.289495. URL http://dl.acm.org/citation.cfm?id=289444.289495. (Cited on pages 1 and 181.)

Carl Gutwin and Saul Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Comput. Supported Coop. Work*, 11(3):411–446, November 2002. ISSN 0925-9724. doi: 10.1023/A:1021271517844. URL http://dx.doi.org/10.1023/A:1021271517844. (Cited on page 7.)

Carl Gutwin, Gwen Stark, and Saul Greenberg. Support for workspace awareness in educational groupware. In *The First International Conference on Computer Support for Collaborative Learning*, CSCL '95, pages 147–156, Hillsdale, NJ, USA, 1995. L. Erlbaum Associates Inc. ISBN 0-8058-2243-7. doi: 10.3115/222020.222126. URL `http://dx.doi.org/10.3115/222020.222126`. (Cited on page 6.)

Edward Twitchell Hall and Edward T Hall. *The hidden dimension*, volume 1990. Anchor Books New York, 1969. (Cited on pages 2, 6 and 181.)

Chris Harrison and Scott E. Hudson. A new angle on cheap LCDs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, page 537, New York, New York, USA, October 2011. ACM Press. ISBN 9781450307161. doi: 10.1145/2047196.2047266. URL `http://dl.acm.org/citation.cfm?id=2047196.2047266`. (Cited on pages 19, 46, 47, 49 and 219.)

Chris Harrison, Desney Tan, and Dan Morris. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753394. URL `http://doi.acm.org/10.1145/1753326.1753394`. (Cited on page 137.)

Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. Omnitouch: Wearable multitouch interaction everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 441–450, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047255. URL `http://doi.acm.org/10.1145/2047196.2047255`. (Cited on page 137.)

Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. On-body interaction: Armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 69–76, New York, NY, USA, 2012a. ACM. ISBN 978-1-4503-1174-8. doi: 10.1145/2148131.2148148. URL `http://doi.acm.org/10.1145/2148131.2148148`. (Cited on page 137.)

Chris Harrison, Munehiko Sato, and Ivan Poupyrev. Capacitive fingerprinting: exploring user differentiation by sensing electrical properties of the human body. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 537–544, New York, NY, USA, 2012b. ACM. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380183. URL `http://doi.acm.org/10.1145/2380116.2380183`. (Cited on pages 23, 26, 27, 37 and 218.)

Graham Healy and Alan F. Smeaton. An outdoor spatially-aware audio playback platform exemplified by a virtual zoo. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 837–840, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-608-3. doi: 10.1145/1631272.1631427. URL `http://doi.acm.org/10.1145/1631272.1631427`. (Cited on pages 139 and 145.)

Felix Heinrichs, Jürgen Steimle, Daniel Schreiber, and Max Mühlhäuser. Letras: an architecture and framework for ubiquitous pen-and-paper interaction. In *Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems - EICS '10*, page 193, New York, New York, USA, June 2010. ACM Press. ISBN 9781450300834. doi: 10.1145/1822018.1822048. URL `http://dl.acm.org/citation.cfm?id=1822018.1822048`. (Cited on page 112.)

Ken Hinckley. Synchronous gestures for multiple persons and computers. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pages 149–158, New York, NY, USA, 2003. ACM. ISBN 1-58113-636-6. doi: 10.1145/964696.964713. URL `http://doi.acm.org/10.1145/964696.964713`. (Cited on pages 93 and 99.)

David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. Paper windows: Interaction techniques for digital paper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 591–599, New York, NY, USA, 2005. ACM. ISBN 1-58113-998-5. doi: 10.1145/1054972.1055054. URL `http://doi.acm.org/10.1145/1054972.1055054`. (Cited on pages 93, 94, 99, 102 and 109.)

David Holman, Jesse Burstyn, Ryan Brotman, Audrey Younkin, and Roel Vertegaal. Flexkit: A rapid prototyping platform for flexible displays. In *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13 Adjunct, pages 17–18, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2406-9. doi: 10.1145/2508468.2514934. URL `http://doi.acm.org/10.1145/2508468.2514934`. (Cited on page 94.)

Christian Holz and Patrick Baudisch. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 581–590, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753413. URL `http://doi.acm.org/10.1145/1753326.1753413`. (Cited on page 32.)

Christian Holz and Patrick Baudisch. Fiberio: A touchscreen that senses fingerprints. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 41–50, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502021. URL `http://doi.acm.org/10.1145/2501988.2502021`. (Cited on pages 35, 37, 61 and 218.)

Christian Holz, Tovi Grossman, George Fitzmaurice, and Anne Agur. Implanted user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 503–512, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2207745. URL `http://doi.acm.org/10.1145/2207676.2207745`. (Cited on page 137.)

Tomoyuki Honda and Kunio Sakamoto. Omni-directional display system for group activity on life review therapy. In *Proceedings of the 9th International Conference on Entertainment Computing*, ICEC'10, pages 452–455, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15398-4, 978-3-642-15398-3. URL `http://dl.acm.org/citation.cfm?id=1881673.1881743`. (Cited on page 48.)

I. P. Howard and W. B. Templeton. *Human Spatial Orientation*. Wiley, London, 1966. (Cited on page 134.)

H Hua and LD Brown. A new collaborative infrastructure: SCAPE. *Virtual Reality 2003*, 3, 2003. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1191136`. (Cited on pages 45 and 49.)

Jochen Huber. *Mobile Interaction with Large Multimedia Information Spaces*. PhD thesis, Technische Universität Darmstadt, December 2012. (Cited on page 96.)

Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Toward more efficient user interfaces for mobile video browsing : An in-depth exploration of the design space. In *Proceedings of the international conference on Multimedia - MM '10*, pages 341–350, New York, New York, USA, October 2010. ACM Press. ISBN 9781605589336. doi: 10.1145/1873951.1873999. URL `http://dl.acm.org/citation.cfm?id=1873951.1873999`. (Cited on page 96.)

Wolfgang Hürst, Georg Götz, and Martina Welte. Interactive video browsing on mobile devices. In *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*, page 247, New York, New York, USA, September 2007. ACM Press. ISBN 9781595937025. doi: 10.1145/1291233.1291284. URL `http://dl.acm.org/citation.cfm?id=1291233.1291284`. (Cited on page 95.)

Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. *Hum.-Comput. Interact.*, 1(4):311–338, December 1985. ISSN 0737-0024. doi: 10.1207/s15327051hci0104_2. URL `http://dx.doi.org/10.1207/s15327051hci0104_2`. (Cited on pages 3, 13 and 134.)

IEEE, editor. *UlteriorScape: Interactive Optical Superimposition on a View-Dependent Tabletop Display*, 2008. IEEE Computer Society Press. ISBN 9781424428984. doi: 10.1109/TABLETOP.2008.4660205. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4660205http://ci.nii.ac.jp/naid/110008722700/`. (Cited on page 48.)

InterSense. InterSense. InterSense IS-900 Precision Motion Tracker. `http://www.isense.com`, 2006. [Online; accessed 17-Dec-2013]. (Cited on page 43.)

Petra Isenberg, Danyel Fisher, Sharoda a Paul, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. Co-located collaborative visual analytics around a tabletop display. *IEEE transactions on visualization and computer graphics*, 18(5):689–702, May 2012. ISSN

1941-0506. doi: 10.1109/TVCG.2011.287. URL http://www.ncbi.nlm.nih.gov/pubmed/22156107. (Cited on pages 2, 7, 18 and 19.)

Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM. ISBN 0-89791-802-9. doi: 10.1145/258549.258715. URL http://doi.acm.org/10.1145/258549.258715. (Cited on page 91.)

A. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. In *2nd Int'l Conference on Audio- and Video-based Biometric Person Authentication, Washington D.C.* 2nd Int'l Conference on Audio- and Video-based Biometric Person Authentication, Washington D.C., 1999. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3468. (Cited on page 32.)

Hao Jiang, Alain Viel, Meekal Bajaj, Robert A. Lue, and Chia Shen. CThru: exploration in a video-centered information space for educational purposes. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, page 1247, New York, New York, USA, April 2009. ACM Press. ISBN 9781605582467. doi: 10.1145/1518701.1518887. URL http://dl.acm.org/citation.cfm?id=1518701.1518887. (Cited on page 96.)

Robert Johansen. *GroupWare: Computer Support for Business Teams*. The Free Press, New York, NY, USA, 1988. ISBN 0029164915. (Cited on page 5.)

Sergi Jordà. The reactable: Tangible and tabletop music performance. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 2989–2994, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: 10.1145/1753846.1753903. URL http://doi.acm.org/10.1145/1753846.1753903. (Cited on pages 141 and 145.)

Y Kakehi, T Hosomi, M Iida, T Naemura, and M Matsushita. Transparent Tabletop Interface for Multiple Users on Lumisight Table. In *Horizontal Interactive Human-Computer Systems, 2006.*, pages 143–150. IEEE, 2006. ISBN 076952494X. doi: 10.1109/TABLETOP.2006.34. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1579206. (Cited on page 47.)

Yasuaki Kakehi, Makoto Iida, Takeshi Naemura, Yoshinari Shirai, Mitsunori Matsushita, and Takeshi Ohguro. Lumisight Table: An Interactive view-Dependent Tabletop Display Surrounded by Multiple Users. *IEEE Computer Graphics and Applications*, 25(1): 48–53, 2005. (Cited on pages 48 and 219.)

Martin Kaltenbrunner. reactivision and tuio: A tangible tabletop toolkit. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 9–16, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-733-2. doi:

10.1145/1731903.1731906. URL `http://doi.acm.org/10.1145/1731903.1731906`. (Cited on page 142.)

Martin Kaltenbrunner and Ross Bencina. reactivision: A computer-vision framework for table-based tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pages 69–74, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1226983. URL `http://doi.acm.org/10.1145/1226969.1226983`. (Cited on page 142.)

Martin Kaltenbrunner, Till Bovermann, Ross Bencina, and Enrico Costanza. Tuio: A protocol for table-top tangible user interfaces. In *In Proceedings of the 2 nd Interactive Sonification Workshop*. Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005), 2005. (Cited on page 58.)

Abhijit Karnik, Diego Martinez Plasencia, Walterio Mayol-Cuevas, and Sriram Subramanian. PiVOT : Personalized View-Overlays for Tabletops. In *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12*, pages 271–280, New York, New York, USA, October 2012. ACM Press. ISBN 9781450315807. doi: 10.1145/2380116.2380151. URL `http://dl.acm.org/citation.cfm?id=2380116.2380151`. (Cited on pages 20, 50, 51, 61 and 219.)

Thorsten Karrer, Moritz Wittenhagen, and Jan Borchers. Pocketdragon: a direct manipulation video navigation interface for mobile devices. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '09*, New York, New York, USA, September 2009. ACM Press. ISBN 9781605582818. doi: 10.1145/1613858.1613917. URL `http://dl.acm.org/citation.cfm?id=1613858.1613917`. (Cited on page 95.)

Thorsten Karrer, Moritz Wittenhagen, Leonhard Lichtschlag, Florian Heller, and Jan Borchers. Pinstripe: Eyes-free continuous input on interactive clothing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1313–1322, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979137. URL `http://doi.acm.org/10.1145/1978942.1979137`. (Cited on page 137.)

Mohammadreza Khalilbeigi, Roman Lissermann, Jan Riemann, Dima Burlak, and Jürgen Steimle. Xpaaand: Interacting with rollable displays. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 307–307, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0399-6. doi: 10.1145/1936652.1936735. URL `http://doi.acm.org/10.1145/1936652.1936735`. (Cited on pages 16 and 186.)

Mohammadreza Khalilbeigi, Roman Lissermann, Max Mühlhäuser, and Jürgen Steimle. Xpaaand: Interaction techniques for rollable displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2729–2732, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979344.

URL http://doi.acm.org/10.1145/1978942.1979344. (Cited on pages 16, 186 and 187.)

Mohammadreza Khalilbeigi, Roman Lissermann, Wolfgang Kleine, and Jürgen Steimle. Foldme: Interacting with double-sided foldable displays. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 33–40, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1174-8. doi: 10.1145/2148131.2148142. URL http://doi.acm.org/10.1145/2148131.2148142. (Cited on pages 16 and 186.)

Sara Kiesler and Jonathon N. Cummings. What do we know about proximity and distance in work groups? a legacy of research, 2002. (Cited on pages 7 and 21.)

Johan Kildal, Susanna Paasovaara, and Viljakaisa Aaltonen. Kinetic device: Designing interactions with a deformable mobile interface. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1871–1876, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1016-1. doi: 10.1145/2212776.2223721. URL http://doi.acm.org/10.1145/2212776.2223721. (Cited on page 186.)

Seokhwan Kim, Xiang Cao, Haimo Zhang, and Desney Tan. Enabling concurrent dual views on common lcd screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2175–2184, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208369. URL http://doi.acm.org/10.1145/2207676.2208369. (Cited on pages 19, 46 and 49.)

Kimoto. Kimoto 100 SXE. http://www.kimoto.co.jp/english/products/light.html, 2012. [Online; accessed 17-Dec-2013]. (Cited on pages 55 and 63.)

David Kirsh. The intelligent use of space. *Artif. Intell.*, 73(1-2):31–68, February 1995. ISSN 0004-3702. doi: 10.1016/0004-3702(94)00017-U. URL http://dx.doi.org/10.1016/0004-3702(94)00017-U. (Cited on pages 3, 11, 87, 88, 92, 104 and 105.)

Yoshifumi Kitamura, Takashige Konishi, Sumihiko Yamamoto, and Fumio Kishino. Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, number August in SIGGRAPH '01, pages 231–240. ACM Press, 2001. ISBN 158113374X. doi: 10.1145/383259.383285. URL http://doi.acm.org/10.1145/383259.383285. (Cited on pages 43, 44, 49 and 219.)

Yoshifumi Kitamura, Wataru Osawa, Tokuo Yamaguchi, Haruo Takemura, and Fumio Kishino. A display table for strategic collaboration preserving private and public information. In *Proceedings of the 4th International Conference on Entertainment Computing*, ICEC'05, pages 167–179, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 3-540-29034-6, 978-3-540-29034-6. doi: 10.1007/11558651_17. URL http://dx.doi.org/10.1007/11558651_17. (Cited on page 43.)

Alexander Kulik, André Kunert, Stephan Beck, Roman Reichel, Roland Blach, Armin Zink, and Bernd Froehlich. C1x6: A Stereoscopic Six-User Display for Co-located Collaboration in Shared Virtual Environments. *ACM Transactions on Graphics (TOG)*, 30(6):188, 2011. doi: 10.1145/2024156.2024222. URL `http://dl.acm.org/citation.cfm?doid=2024156.2024222`. (Cited on pages 40, 49 and 62.)

Fusako Kusunoki, Ikuko Eguchi Yairi, and Takuichi NIshimura. Multi-audible table for collaborative work. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, pages 67–73, New York, NY, USA, 2004. ACM. ISBN 1-58113-882-2. doi: 10.1145/1067343.1067351. URL `http://doi.acm.org/10.1145/1067343.1067351`. (Cited on pages 141 and 145.)

Byron Lahey, Audrey Girouard, Winslow Burleson, and Roel Vertegaal. Paperphone: Understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1303–1312, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979136. URL `http://doi.acm.org/10.1145/1978942.1979136`. (Cited on page 186.)

Kevin A. Li, Patrick Baudisch, and Ken Hinckley. Blindsight: Eyes-free access to mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1389–1398, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357273. URL `http://doi.acm.org/10.1145/1357054.1357273`. (Cited on page 138.)

Peter Lincoln, Andrew Nashel, Adrian Ilie, Herman Towles, Gregory Welch, and Henry Fuchs. Multi-view lenticular display for group teleconferencing. In *Proceedings of the 2Nd International Conference on Immersive Telecommunications*, IMMERSCOM '09, pages 22:1–22:8, ICST, Brussels, Belgium, Belgium, 2009. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). ISBN 978-963-9799-39-4. URL `http://dl.acm.org/citation.cfm?id=1594108.1594137`. (Cited on pages 44 and 49.)

Roman Lissermann, Simon Olberding, Max Mühlhäuser, and Jürgen Steimle. Interacting with videos on paper-like displays. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 2579–2584, New York, NY, USA, 2012a. ACM. ISBN 978-1-4503-1016-1. doi: 10.1145/2212776.2223839. URL `http://doi.acm.org/10.1145/2212776.2223839`. (Cited on pages 15, 16 and 87.)

Roman Lissermann, Simon Olberding, Benjamin Petry, Max Mühlhäuser, and Jürgen Steimle. Papervideo: Interacting with videos on multiple paper-like displays. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 129–138, New York, NY, USA, 2012b. ACM. ISBN 978-1-4503-1089-5. doi: 10.1145/2393347.2393372. URL `http://doi.acm.org/10.1145/2393347.2393372`. (Cited on pages 15, 87 and 136.)

Roman Lissermann, Benjamin Petry, Mohammadreza Khalilbeigi, and Jochen Huber. Toward mobile video interaction with rollable displays. In *Proceedings of Beyond Flat Displays: Towards Shaped and Deformable Interactive Surfaces, in conjunction with ITS 2012*, November 2012c. (Cited on page 187.)

Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. Earput: Augmenting behind-the-ear devices for ear-based interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1323–1328, New York, NY, USA, 2013a. ACM. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2468592. URL http://doi.acm.org/10.1145/2468356.2468592. (Cited on pages 15 and 136.)

Roman Lissermann, Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Permulin: Collaboration on interactive surfaces with personal in- and output. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1533–1538, New York, NY, USA, 2013b. ACM. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2468630. URL http://doi.acm.org/10.1145/2468356.2468630. (Cited on pages 15 and 18.)

Roman Lissermann, Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Permulin: Personal in- and output on interactive surfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 3083–3086, New York, NY, USA, 2013c. ACM. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2479616. URL http://doi.acm.org/10.1145/2468356.2479616. (Cited on pages 15 and 18.)

Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. Earput: Augmenting ear-worn devices for ear-based interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI '14, page (in press). ACM, 2014a. (Cited on pages 16 and 136.)

Roman Lissermann, Jochen Huber, Martin Schmitz, Jürgen Steimle, and Max Mühlhäuser. Permulin: Mixed-focus collaboration on multi-view tabletops. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3191–3200, New York, NY, USA, 2014b. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557405. URL http://doi.acm.org/10.1145/2556288.2557405. (Cited on pages 15 and 18.)

P Luff and C Heath. Mobility in collaboration. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, CSCW '98, pages 305–314. ACM, 1998. ISBN 1581130090. doi: 10.1145/289444.289505. URL http://portal.acm.org/citation.cfm?doid=289444.289505. (Cited on page 92.)

W. Mackay and D. Pagani. Video mosaic: Laying out time in a physical space. In *Proceedings of the Second ACM International Conference on Multimedia*, MULTIMEDIA '94, pages 165–172, New York, NY, USA, 1994. ACM. ISBN 0-89791-686-7. doi: 10.1145/192593.192646. URL http://doi.acm.org/10.1145/192593.192646. (Cited on pages 11, 88 and 96.)

Thomas W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, January 1983. ISSN 1046-8188. doi: 10.1145/357423.357430. URL http://doi.acm.org/10.1145/357423.357430. (Cited on page 105.)

Richard Mander, Gitta Salomon, and Yin Yin Wong. A “pile” metaphor for supporting casual organization of information. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, pages 627–634, New York, New York, USA, June 1992. ACM Press. ISBN 0897915135. doi: 10.1145/142750.143055. URL http://dl.acm.org/citation.cfm?id=142750.143055. (Cited on page 105.)

Knut Manske, Max Mühlhäuser, Simon Vogl, and Morris Goldberg. Obvi: Hierarchical 3d video-browsing. In *Proceedings of the 6th ACM International Multimedia Conference, Bristol, UK, Nov. 1998*, pages 369–374. ACM, ACM Press, New York, NY, November 1998. ISBN 1-58113-036-8. (Cited on page 96.)

Nicolai Marquardt, Johannes Kiemer, and Saul Greenberg. What caused that touch?: Expressive interaction with a surface through fiduciary-tagged gloves. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 139–142, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0399-6. doi: 10.1145/1936652.1936680. URL http://doi.acm.org/10.1145/1936652.1936680. (Cited on pages 31 and 37.)

Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. Cross-device interaction via micro-mobility and f-formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 13–22, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380121. URL http://doi.acm.org/10.1145/2380116.2380121. (Cited on page 93.)

Mitsunori Matsushita, Makoto Iida, Takeshi Ohguro, Yoshinari Shirai, Yasuaki Kakehi, and Takeshi Naemura. Lumisight Table:A face-to-face collaboration support system that optimizes direction of projected information to each stakeholder. In *CSCW2004*, pages 274–283. ACM Press, 2004. ISBN 1581138105. doi: 10.1145/1031607.1031653. URL http://dl.acm.org/citation.cfm?id=1031607.1031653. (Cited on pages 20, 47 and 49.)

Wojciech Matusik, Clifton Forlines, and Hanspeter Pfister. Multiview user interfaces with an automultiscopic display. In *Proceedings of the working conference on Advanced visual interfaces - AVI '08*, page 363, New York, New York, USA, May 2008. ACM Press. ISBN 9781605581415. doi: 10.1145/1385569.1385634. URL http://dl.acm.org/citation.cfm?id=1385569.1385634. (Cited on pages 19, 44 and 49.)

David Merrill, Jeevan Kalanithi, and Pattie Maes. Siftables: Towards sensor network user interfaces. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pages 75–78, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1226984. URL http://doi.acm.org/10.1145/1226969.1226984. (Cited on pages 93 and 99.)

Robert Mertens, Holger Schneider, Olaf Müller, and Oliver Vornberger. Hypermedia navigation concepts for lecture recordings. In *Proceedings of E-Learn*, pages 2480–2847. AACE, 2004. (Cited on page 95.)

Christian Metzger, Matt Anderson, and Thad Starner. Freedigiter: A contact-free device for gesture control. In *Proceedings of the Eighth International Symposium on Wearable Computers*, ISWC '04, pages 18–21, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2186-X. doi: 10.1109/ISWC.2004.23. URL http://dx.doi.org/10.1109/ISWC.2004.23. (Cited on pages 138 and 153.)

Tobias Meyer and Dominik Schmidt. Idwristbands: Ir-based user identification on multi-touch surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 277–278, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0399-6. doi: 10.1145/1936652.1936714. URL http://doi.acm.org/10.1145/1936652.1936714. (Cited on pages 30 and 37.)

Pranav Mistry. ThirdEye: a technique that enables multiple viewers to see different content on a single display screen. In *ACM SIGGRAPH ASIA 2009 Posters (SIGGRAPH ASIA '09)*, pages 1–1, New York, New York, USA, December 2009. ACM Press. doi: 10.1145/1666778.1666807. URL http://dl.acm.org/citation.cfm?id=1666778.1666807http://fluid.media.mit.edu/people/pranav/current/thirdeye.html. (Cited on pages 19, 40 and 49.)

Pranav Mistry and Pattie Maes. Sixthsense: A wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, SIGGRAPH ASIA '09, pages 11:1–11:1, New York, NY, USA, 2009. ACM. doi: 10.1145/1667146.1667160. URL http://doi.acm.org/10.1145/1667146.1667160. (Cited on page 137.)

MIT. OpenCourseWare. http://www.apple.com/education/ipad/itunes-u/, 2002. [Online; accessed 11-Mar-2014]. (Cited on page 106.)

A. Montagu. *Touching: The Human Significance of the Skin*. Perennial Library. Perennial Library, 1986. ISBN 9780060960285. URL http://books.google.de/books?id=I48QAQAAIAAJ. (Cited on pages 138 and 186.)

Meredith Ringel Morris, Dan Morris, and Terry Winograd. Individual audio channels with single display groupware: Effects on communication and task strategy. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 242–251, New York, NY, USA, 2004. ACM. ISBN 1-58113-810-5. doi:

10.1145/1031607.1031646. URL http://doi.acm.org/10.1145/1031607.1031646. (Cited on pages 141, 145, 168, 173 and 221.)

Meredith Ringel Morris, Anthony Cassanego, Andreas Paepcke, Terry Winograd, Anne Marie Piper, and Anqi Huang. Mediating group dynamics through tabletop interface design. *IEEE Computer Graphics and Applications*, 26(5):65–73, 2006. ISSN 02721716. doi: 10.1109/MCG.2006.114. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1683695. (Cited on pages 8 and 22.)

Christian Müller-Tomfelde. *Tabletops - Horizontal Interactive Displays*. Springer Publishing Company, Incorporated, 1st edition, 2010. ISBN 1849961123, 9781849961127. (Cited on page 21.)

Brad A. Myers, Herb Stiel, and Robert Gargiulo. Collaboration using multiple pdas connected to a pc. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98, pages 285–294, New York, NY, USA, 1998. ACM. ISBN 1-58113-009-0. doi: 10.1145/289444.289503. URL http://doi.acm.org/10.1145/289444.289503. (Cited on page 29.)

Y. Nakamura, T. Nishimura, H. Itoh, and H. Nakashima. Id-cobit: a battery-less information terminal with data upload capability. In *Industrial Electronics Society, 2003. IECON '03. The 29th Annual Conference of the IEEE*, volume 3, pages 2511–2516 Vol.3. IEEE, Nov 2003. doi: 10.1109/IECON.2003.1280640. (Cited on page 141.)

Donald A. Norman and Stephen W. Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1986. ISBN 0898597811. (Cited on page 15.)

Cenker Oden, Aytul Ercil, and Burak Buke. Combining implicit polynomials and geometric features for hand recognition. *Pattern Recognition Letters*, 24 (13):2145–2152, September 2003. ISSN 01678655. doi: 10.1016/S0167-8655(03)00087-4. URL http://linkinghub.elsevier.com/retrieve/pii/S0167865503000874. (Cited on page 32.)

Kenton O'Hara and Abigail Sellen. A comparison of reading paper and on-line documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, pages 335–342, New York, New York, USA, March 1997. ACM Press. ISBN 0897918029. doi: 10.1145/258549.258787. URL http://dl.acm.org/citation.cfm?id=258549.258787. (Cited on pages 87 and 120.)

OpenAL. OpenAL. http://connect.creativelabs.com/openal/default.aspx, 2010. [Online; accessed 11-Mar-2014]. (Cited on pages 113 and 162.)

OpenNI. OpenNI. http://www.openni.org/, 2011. [Online; accessed 11-Mar-2014]. (Cited on page 56.)

OptiTrack. OptiTrack. http://www.naturalpoint.com/optitrack/, 2008. [Online; accessed 11-Mar-2014]. (Cited on page 112.)

Robert J. Orr and Gregory D. Abowd. The smart floor: A mechanism for natural user identification and tracking. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, pages 275–276, New York, NY, USA, 2000. ACM. ISBN 1-58113-248-4. doi: 10.1145/633292.633453. URL http://doi.acm.org/10.1145/633292.633453. (Cited on pages 31, 34, 37 and 61.)

OSC-protocol. OSC-Protocol . http://opensoundcontrol.org/, 2014. [Online; accessed 03-June-2014]. (Cited on page 56.)

Ken Perlin, Salvatore Paxia, and Joel S. Kollin. An autostereoscopic display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 319–326, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN 1-58113-208-5. doi: 10.1145/344779.344933. URL http://dx.doi.org/10.1145/344779.344933. (Cited on pages 43, 49 and 55.)

Tom Peterka, Robert Kooima, Javier Girado, Jinghua Ge, Daniel J. Sandin, Andrew E. Johnson, Jason Leigh, Jürgen P. Schulze, and Thomas A. DeFanti. Dynallax: Solid state dynamic parallax barrier autostereoscopic vr display. In William R. Sherman, Ming Lin, and Anthony Steed, editors, *VR*, pages 155–162. IEEE Computer Society, 2007. ISBN 1-4244-0905-5. (Cited on pages 43 and 49.)

Raf Ramakers, Davy Vanacken, Kris Luyten, Karin Coninx, and Johannes Schöning. Carpus: A non-intrusive user identification technique for interactive surfaces. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 35–44, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380123. URL http://doi.acm.org/10.1145/2380116.2380123. (Cited on pages 23, 33, 37, 60 and 61.)

Jun Rekimoto. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proceedings of the 10th annual ACM symposium on User interface software and technology (UIST '97)*, pages 31–39. ACM, 1997. ISBN 0897918819. doi: 10.1145/263407.263505. URL http://portal.acm.org/citation.cfm?id=263505. (Cited on page 109.)

Jun Rekimoto and Masanori Saitoh. Augmented surfaces: A spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 378–385, New York, NY, USA, 1999. ACM. ISBN 0-201-48559-1. doi: 10.1145/302979.303113. URL http://doi.acm.org/10.1145/302979.303113. (Cited on page 22.)

Jun Rekimoto, Brygg Ullmer, and Haruo Oba. DataTiles: a modular platform for mixed physical and graphical interactions. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*, pages 269–276, New York, New York, USA, March 2001. ACM Press. ISBN 1581133278. doi: 10.1145/365024.365115. URL `http://dl.acm.org/citation.cfm?id=365024.365115`. (Cited on pages 93 and 99.)

Stephan Richter, Christian Holz, and Patrick Baudisch. Bootstrapper: Recognizing tabletop users by their shoes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1249–1252, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208577. URL `http://doi.acm.org/10.1145/2207676.2208577`. (Cited on pages 34, 37 and 218.)

Jan Riemann. Multiprojektorsystem zur simulation hochauflösender papierähnlicher displays. Bachelor's thesis, TU Darmstadt, 2010. (Cited on page 112.)

Mike Rooke and Roel Vertegaal. Physics on display: Tangible graphics on hexagonal bezel-less screens. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '10, pages 233–236, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-841-4. doi: 10.1145/1709886.1709932. URL `http://doi.acm.org/10.1145/1709886.1709932`. (Cited on pages 93 and 99.)

Volker Roth, Philipp Schmidt, and Benjamin Güldenring. The ir ring: Authenticating users' touches on a multi-touch display. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 259–262, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0271-5. doi: 10.1145/1866029.1866071. URL `http://doi.acm.org/10.1145/1866029.1866071`. (Cited on pages 30, 37 and 61.)

Kathy Ryall, Clifton Forlines, Chia Shen, and Meredith Ringel Morris. Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 284–293, New York, NY, USA, 2004. ACM. ISBN 1-58113-810-5. doi: 10.1145/1031607.1031654. URL `http://doi.acm.org/10.1145/1031607.1031654`. (Cited on page 22.)

Sakurai Sakurai, Yoshifumi Kitamura, Sriram Subramanian, and Fumio Kishino. Visibility control using revolving polarizer. In *3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008.*, pages 161–168. IEEE, October 2008. ISBN 978-1-4244-2897-7. doi: 10.1109/TABLETOP.2008.4660200. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4660200`. (Cited on pages 41, 42, 49 and 219.)

Satoshi Sakurai, Yoshifumi Kitamura, Sriram Subramanian, and Fumio Kishino. A visibility control system for collaborative digital table. *Personal and Ubiquitous Computing*, 13 (8):619–632, July 2009. ISSN 1617-4909. URL `http://dl.acm.org/citation.cfm?id=1644246.1644261`. (Cited on page 41.)

Samsung. Samsung amazing flexible display [ces 2013]. https://www.youtube.com/watch?v=pU0hb-2Srj4, 2013. [Online; accessed 11-Mar-2014]. (Cited on page 186.)

R Sanchez-Reillo. Biometric identification through hand geometry measurements. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1168–1171, 2000. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=879796. (Cited on page 32.)

Munehiko Sato, Ivan Poupyrev, and Chris Harrison. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 483–492, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2207743. URL http://doi.acm.org/10.1145/2207676.2207743. (Cited on pages 26 and 218.)

Chris Schmandt. Audio hallway: A virtual acoustic environment for browsing. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, pages 163–170, New York, NY, USA, 1998. ACM. ISBN 1-58113-034-1. doi: 10.1145/288392.288597. URL http://doi.acm.org/10.1145/288392.288597. (Cited on pages 139 and 145.)

Chris Schmandt and Atty Mullins. Audiostreamer: Exploiting simultaneity for listening. In *Conference Companion on Human Factors in Computing Systems*, CHI '95, pages 218–219, New York, NY, USA, 1995. ACM. ISBN 0-89791-755-3. doi: 10.1145/223355.223533. URL http://doi.acm.org/10.1145/223355.223533. (Cited on pages 139 and 145.)

Dominik Schmidt, Ming Ki Chong, and Hans Gellersen. Handsdown: Hand-contour-based user identification for interactive surfaces. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 432–441, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-934-3. doi: 10.1145/1868914.1868964. URL http://doi.acm.org/10.1145/1868914.1868964. (Cited on pages 23, 33 and 37.)

Kjeld Schmidt and Carla Simonee. Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)*, 5 (2-3):155–200, 1996. ISSN 0925-9724. doi: 10.1007/BF00133655. URL http://link.springer.com/10.1007/BF00133655. (Cited on page 7.)

Klaus Schoeffmann and Manfred del Fabro. Hierarchical video browsing with a 3D carousel. In *Proceedings of the 19th ACM international conference on Multimedia - MM '11*, page 827, New York, New York, USA, November 2011. ACM Press. ISBN 9781450306164. doi: 10.1145/2072298.2072479. URL http://dl.acm.org/citation.cfm?id=2072298.2072479. (Cited on page 95.)

Klaus Schoeffmann, Frank Hopfgartner, Oge Marques, Laszlo Böszörmenyi, and Joemon Jose. Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1):018004, mar 2010. doi: 10.1117/6.0000005. URL `http://link.aip.org/link/?SV2/1/018004/1`. (Cited on page 95.)

J. Schöning, M. Rohs, and A. Krüger. Using Mobile Phones to Spontaneously Authenticate and Interact withMulti-touch Surfaces. In *PPD'08. Workshop on designing multi-touch interaction techniques for coupled public and private displays*. PPD'08. Workshop on designing multi-touch interaction techniques for coupled public and private displays, 2008. (Cited on pages 30, 37 and 218.)

Julia Schwarz, Chris Harrison, Scott Hudson, and Jennifer Mankoff. Cord input: An intuitive, high-accuracy, multi-degree-of-freedom input method for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1657–1660, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753573. URL `http://doi.acm.org/10.1145/1753326.1753573`. (Cited on page 137.)

Carsten Schwesig, Ivan Poupyrev, and Eijiro Mori. Gummi: A bendable computer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 263–270, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: 10.1145/985692.985726. URL `http://doi.acm.org/10.1145/985692.985726`. (Cited on page 186.)

Stacey D Scott. *Territoriality in Collaborative Tabletop Workspaces*. PhD thesis, The University of Calgary, Calgary, Alberta, 2005. (Cited on page 21.)

Stacey D Scott, M Sheelagh, T Carpendale, and Kori M Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work CSCW 04*, CSCW '04, pages 294 – 303. ACM Press, 2004. ISBN 1581138105. doi: 10.1145/1031607.1031655. URL `http://portal.acm.org/citation.cfm?doid=1031607.1031655`. (Cited on pages 2, 8, 22, 38, 78, 92, 120, 127, 169, 172 and 181.)

Julian Seifert, Adalberto Simeone, Dominik Schmidt, Paul Holleis, Christian Reinartz, Matthias Wagner, Hans Gellersen, and Enrico Rukzio. Mobisurf: Improving co-located collaboration through integrating mobile devices and interactive surfaces. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, pages 51–60, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1209-7. doi: 10.1145/2396636.2396644. URL `http://doi.acm.org/10.1145/2396636.2396644`. (Cited on pages 22 and 36.)

Abigail J. Sellen and Richard H. R. Harper. *The Myth of the Paperless Office*. The MIT Press, 2001. ISBN 0262194643. URL `http://www.amazon.com/Myth-Paperless-Office-Abigail-Sellen/dp/0262194643`. (Cited on pages 3, 11, 87, 88, 92 and 161.)

Freescale Semiconductor. Proximity capacitive touch sensor controller: Mpr121, 2010. (Cited on pages 149 and 154.)

Sherrington CS. On the proprioceptive system, especially in its reflex aspect. In *Brain*. Brain, 1906. (Cited on pages 138 and 147.)

Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley, 2004. ISBN 0321197860. (Cited on page 72.)

Garth B. D. Shoemaker and Kori M. Inkpen. Single display privacyware. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*, pages 522–529, New York, New York, USA, March 2001. ACM Press. ISBN 1581133278. doi: 10.1145/365024.365349. URL `http://dl.acm.org/citation.cfm?id=365024.365349`. (Cited on pages 20, 39, 40, 49 and 219.)

David Small and Hiroshi Ishii. Design of spatially aware graspable displays. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pages 367–368, New York, NY, USA, 1997. ACM. ISBN 0-89791-926-2. doi: 10.1145/1120212.1120437. URL `http://doi.acm.org/10.1145/1120212.1120437`. (Cited on pages 90 and 99.)

Ross T. Smith and Wayne Piekarski. Public and private workspaces on tabletop displays. In *Proceedings of the Ninth Conference on Australasian User Interface - Volume 76*, AUIC '08, pages 51–54, Darlinghurst, Australia, Australia, 2008. Australian Computer Society, Inc. ISBN 978-1-920682-57-6. URL `http://dl.acm.org/citation.cfm?id=1378337.1378347`. (Cited on pages 44 and 49.)

Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr. Aireal: Interactive tactile experiences in free air. *ACM Trans. Graph.*, 32(4):134:1–134:10, July 2013. ISSN 0730-0301. doi: 10.1145/2461912.2462007. URL `http://doi.acm.org/10.1145/2461912.2462007`. (Cited on page 187.)

Tomas Sokoler and H\aa kan Edeholt. Physically embodied video snippets supporting collaborative exploration of video material during design sessions. In *Proceedings of the second Nordic conference on Human-computer interaction - NordiCHI '02*, page 139, New York, New York, USA, October 2002. ACM Press. ISBN 1581136161. doi: 10.1145/572020.572037. URL `http://dl.acm.org/citation.cfm?id=572020.572037`. (Cited on page 97.)

Sony. Sony develops a rollable OLED screen. `https://www.youtube.com/watch?v=9OvTLg4i2_U`, 2010. [Online; accessed 11-Mar-2014]. (Cited on page 186.)

Martin Spindler, Sophie Stellmach, and Raimund Dachselt. PaperLens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 69–76. ACM, 2009. ISBN

9781605587332. doi: 10.1145/1731903.1731920. URL `http://portal.acm.org/citation.cfm?id=1731903.1731920`. (Cited on pages 93, 99, 100 and 102.)

Martin Spindler, Christian Tominski, Heidrun Schumann, and Raimund Dachselt. Tangible views for information visualization. In Antonio Krüger, Johannes Schöning, Daniel Wigdor, and Michael Haller, editors, *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, ITS '10, pages 157–166. ACM, 2010. ISBN 9781450303996. doi: 10.1145/1936652.1936684. URL `http://portal.acm.org/citation.cfm?id=1936684`. (Cited on pages 93, 94, 99, 100 and 102.)

Martin Spindler, Wolfgang Büschel, and Raimund Dachselt. Use your head: Tangible windows for 3d information spaces in a tabletop environment. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, pages 245–254, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1209-7. doi: 10.1145/2396636.2396674. URL `http://doi.acm.org/10.1145/2396636.2396674`. (Cited on page 100.)

Martin Spindler, Wolfgang Büschel, Charlotte Winkler, and Raimund Dachselt. Tangible displays for the masses: Spatial interaction with handheld displays by using consumer depth cameras. *Personal Ubiquitous Comput.*, 18(5):1213–1225, June 2014. ISSN 1617-4909. doi: 10.1007/s00779-013-0730-7. URL `http://dx.doi.org/10.1007/s00779-013-0730-7`. (Cited on pages 93 and 102.)

Jürgen Steimle and Simon Olberding. When mobile phones expand into handheld tabletops. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 271–280, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1016-1. doi: 10.1145/2212776.2212805. URL `http://doi.acm.org/10.1145/2212776.2212805`. (Cited on page 186.)

Jürgen Steimle, Roman Lissermann, Simon Olberding, Mohammadreza Khalibeigi, Wolfgang Kleine, and Max Mühlhäuser. Be-greifbare interaktionen mit größenveränderbaren bildschirmen. *i-com, Oldenbourg Verlag*, Vol. 11(2):26–31, February 2012. (Cited on page 16.)

Jürgen Steimle, Andreas Jordt, and Pattie Maes. Flexpad: Highly flexible bending interactions for projected handheld displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 237–246, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2470688. URL `http://doi.acm.org/10.1145/2470654.2470688`. (Cited on page 186.)

Jason Stewart, BB Bederson, and Allison Druin. Single display groupware: a model for co-present collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 14(98), 1999. URL `http://dl.acm.org/citation.cfm?id=303064`. (Cited on page 30.)

Anselm Strauss and Juliet Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, volume 3. Sage, 2008. ISBN 0803959400. URL http://emedien.sub.uni-hamburg.de/han/OvidDatenbankenPsychologie/kataloge.uni-hamburg.de:80/DB=1/SET=1/TTL=1/SHW?FRST=1. (Cited on pages 120, 127, 152, 163, 167 and 175.)

Ellen A. Strommen. *Verbal Self-Regulation in a Children's Game: Impulsive Errors on "Simon Says"*, volume 4. Wiley, 1973. (Cited on page 159.)

Ramona E. Su and Brian P. Bailey. Put them where? towards guidelines for positioning large displays in interactive workspaces. In *Proc. of Graphics Interface '05*, page 337349, 2005. (Cited on page 22.)

Masanori Sugimoto, Kazuhiro Hosoi, and Hiromichi Hashizume. Caretta: A system for supporting face-to-face collaboration by integrating personal and shared spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 41–48, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: 10.1145/985692.985698. URL http://doi.acm.org/10.1145/985692.985698. (Cited on pages 8, 22 and 36.)

Atsushi Sugiura and Yoshiyuki Koseki. A user interface using fingerprint recognition: Holding commands and data objects on fingers. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, pages 71–79, New York, NY, USA, 1998. ACM. ISBN 1-58113-034-1. doi: 10.1145/288392.288575. URL http://doi.acm.org/10.1145/288392.288575. (Cited on page 32.)

Qibin Sun and Wolfgang Hurst. Video browsing on handheld devices - interface designs for the next generation of mobile video players. *IEEE Multimedia*, 15(3):76–83, 2008. ISSN 1070-986X. doi: http://doi.ieeecomputersociety.org/10.1109/MMUL.2008.66. (Cited on page 96.)

Desney S. Tan and Mary Czerwinski. Effects of visual separation and physical discontinuities when distributing information across multiple displays. In *In Proceedings of Interact 2003*, pages 252–255, 2003. (Cited on page 22.)

CA Tanase, RD Vatavu, SG Pentiuc, and Adrian Graur. Detecting and Tracking Multiple Users in the Proximity of Interactive Tabletops. *Advances in Electrical and Computer Engineering*, 8(2), 2008. URL http://www.doaj.org/doaj?func=fulltext&aId=494856. (Cited on pages 27, 28, 37 and 218.)

Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, volume pp of *CHI '06*, pages 1181–1190, New York, New York, USA, 2006. ACM Press. ISBN 1595933727. doi:

10.1145/1124772.1124950. URL `http://portal.acm.org/citation.cfm?doid=1124772.1124950`. (Cited on pages 2, 6, 7, 19, 52, 54, 67 and 71.)

Aneesh P. Tarun, Peng Wang, Audrey Girouard, Paul Strohmeier, Derek Reilly, and Roel Vertegaal. Papertab: An electronic paper computer with multiple large flexible electrophoretic displays. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 3131–3134, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2479628. URL `http://doi.acm.org/10.1145/2468356.2479628`. (Cited on pages 87, 94, 95, 97, 99, 102, 186 and 220.)

Lucia Terrenghi, David Kirk, Abigail Sellen, and Shahram Izadi. Affordances for manipulation of physical versus digital media on interactive surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1157–1166, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240799. URL `http://doi.acm.org/10.1145/1240624.1240799`. (Cited on pages 92 and 186.)

Edward Tse, Jonathan Histon, Stacey D. Scott, and Saul Greenberg. Avoiding interference: How people use spatial separation and partitioning in SDG workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, pages 252–261, New York, New York, USA, November 2004. ACM Press. ISBN 1581138105. doi: 10.1145/1031607.1031647. URL `http://dl.acm.org/citation.cfm?id=1031607.1031647`. (Cited on pages 2, 22, 63, 92 and 181.)

B.D. Van Veen and K.M. Buckley. Beamforming: a versatile approach to spatial filtering. *ASSP Magazine, IEEE*, 5(2):4–24, April 1988. ISSN 0740-7467. doi: 10.1109/53.665. (Cited on pages 142 and 145.)

Yolanda. Vazquez-alvarez. *An Investigation of Eyes-Free Spatial Auditory Interfaces for Mobile Devices: Supporting Multitasking and Location-Based Information*. PhD thesis, University of Glasgow, 2013. (Cited on page 142.)

Yolanda Vazquez Alvarez and Stephen A. Brewster. Designing spatial audio interfaces to support multiple audio streams. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, pages 253–256, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-835-3. doi: 10.1145/1851600.1851642. URL `http://doi.acm.org/10.1145/1851600.1851642`. (Cited on page 140.)

Yolanda Vazquez-Alvarez and Stephen A. Brewster. Eyes-free multitasking: The effect of cognitive load on mobile spatial audio interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2173–2176, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979258. URL `http://doi.acm.org/10.1145/1978942.1979258`. (Cited on page 140.)

Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. Auditory display design for exploration in mobile audio-augmented reality. *Personal Ubiquitous Comput.*, 16(8):987–999, December 2012. ISSN 1617-4909. doi: 10.1007/s00779-011-0459-0. URL http://dx.doi.org/10.1007/s00779-011-0459-0. (Cited on pages 140 and 145.)

Polymer Vision. Polymer Vision rollable 6-inch SVGA display. https://www.youtube.com/watch?v=xxhCiLvi5LI, 2011. [Online; accessed 11-Mar-2014]. (Cited on page 186.)

Julie Wagner, Mathieu Nancel, Sean G. Gustafson, Stephane Huot, and Wendy E. Mackay. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1299–1308, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2466170. URL http://doi.acm.org/10.1145/2470654.2466170. (Cited on page 137.)

Ashley Walker, Stephen Brewster, David McGookin, and Adrian Ng. Diary in the sky: A spatial audio display for a mobile calendar. In Ann Blandford, Jean Vanderdonckt, and Phil Gray, editors, *People and Computers XV—Interaction without Frontiers*, pages 531–539. Springer London, 2001. ISBN 978-1-85233-515-1. doi: 10.1007/978-1-4471-0353-0_33. URL http://dx.doi.org/10.1007/978-1-4471-0353-0_33. (Cited on pages 134, 139, 140 and 221.)

James R. Wallace, Stacey D. Scott, Taryn Stutz, Tricia Enns, and Kori Inkpen. Investigating teamwork and taskwork in single- and multi-display groupware systems. *Personal Ubiquitous Comput.*, 13(8):569–581, November 2009. ISSN 1617-4909. doi: 10.1007/s00779-009-0241-8. URL http://dx.doi.org/10.1007/s00779-009-0241-8. (Cited on page 22.)

B Walther-Franks and L Schwarten. User detection for a multi-touch table via proximity sensors. *IEEE Tabletops and Interactive Surfaces*, 1, 2008. URL http://www.tzi.de/fileadmin/user_upload/materials/research/MRT/User_Detection_for_a_Multi-touch_Table_via_Proximity_Sensors.pdf. (Cited on pages 28 and 37.)

Martin Weigel, Vikram Mehta, and Jürgen Steimle. More than touch: Understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 179–188, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557239. URL http://doi.acm.org/10.1145/2556288.2557239. (Cited on page 138.)

Pierre Wellner. The digitaldesk calculator: Tangible manipulation on a desk top display. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*, UIST '91, pages 27–33, New York, NY, USA, 1991. ACM. ISBN 0-89791-451-1.

doi: 10.1145/120782.120785. URL http://doi.acm.org/10.1145/120782.120785. (Cited on pages 21 and 93.)

Daniel Wigdor, Hao Jiang, Clifton Forlines, Michelle Borkin, and Chia Shen. WeSpace: the design development and deployment of a walk-up and share multi-surface visual collaboration system. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1237, New York, New York, USA, 2009. ACM Press. ISBN 9781605582467. doi: 10.1145/1518701.1518886. URL http://dl.acm.org/citation.cfm?id=1518886http://dl.acm.org/citation.cfm?doid=1518701.1518886. (Cited on pages 22 and 36.)

Xiaolin Wu and Guangtao Zhai. Temporal Psychovisual Modulation: A New Paradigm of Information Display. *Signal Processing Magazine, IEEE*, 1(january):136–141, 2013. URL http://arxiv.org/abs/1201.1601http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6375944. (Cited on pages 40, 49, 55, 62 and 64.)

Yan Xiao. Artifacts and collaborative work in healthcare: Methodological, theoretical, and technological implications of the tangible. *J. of Biomedical Informatics*, 38(1):26–33, February 2005. ISSN 1532-0464. doi: 10.1016/j.jbi.2004.11.004. URL http://dx.doi.org/10.1016/j.jbi.2004.11.004. (Cited on pages 7 and 21.)

Lining Yao, Ryuma Niiyama, Jifei Ou, Sean Follmer, Clark Della Silva, and Hiroshi Ishii. Pneui: Pneumatically actuated soft composite materials for shape changing interfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 13–22, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502037. URL http://doi.acm.org/10.1145/2501988.2502037. (Cited on page 186.)

G Ye and H Fuchs. A practical multi-viewer tabletop autostereoscopic display. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium*, pages 147–156. IEEE, 2010. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5643563. (Cited on pages 44 and 49.)

Bo Yi, Xiang Cao, Morten Fjeld, and Shengdong Zhao. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2789–2792, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208678. URL http://doi.acm.org/10.1145/2207676.2208678. (Cited on page 138.)

Ana Zanella and Saul Greenberg. Reducing interference in single display groupware through transparency. In *Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work (ECSCW'01)*, pages 339–358, 2001. URL http://www.springerlink.com/index/K4J6362218T31377.pdf. (Cited on page 8.)

Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. Earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1395–1404, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240836. URL `http://doi.acm.org/10.1145/1240624.1240836`. (Cited on pages 139 and 145.)

Jamie Zigelbaum, Michael S Horn, and Robert J K Jacob. The tangible video editor : Collaborative video editing with active tokens. In *Proceedings of the 1st international conference on Tangible and embedded interaction (TEI '07)*, pages 43 – 46. ACM Press, 2007. ISBN 9781595936196. doi: 10.1145/1226969.1226978. URL `http://www.cs.tufts.edu/jacob/papers/tei07.zigelbaum.pdf`. (Cited on pages 96 and 97.)

List of Figures

List of Tables

CV

Mai 2010 - Sep. 2014	PhD student in computer science *TU Darmstadt, Germany* Research Area: Human-Computer Interaction Advisors: Prof. Dr. Mühlhäuser (TU Darmstadt, Germany) Prof. Dr. Suranga Nanayakkara (SUTD, Singapore) Topic: Novel Devices and Interaction Concepts for Fluid Collaboration
Juni 2014 - Sep. 2014	Intern Scientist *Yahoo Labs, Sunnyvale, USA*
2007 - 2010	M.Sc. in computer science *TU Darmstadt, Germany* Minor: Business
July 2008 - Mai 2009	Exchange program *NTU and NBS, Singapur* Graduate school in computer science, Undergraduate studies in business
2004 - 2007	B.Sc. in computer science *TU Darmstadt, Germany*
2004	Abitur *Heinrich Kleyer Schule, Frankfurt, Germany*